

; Seq. Id. Nos. 3 and 11 from USSN 09/770,693 joined
; in the order SID 11 followed by SID 3
693_11_3 SEQ ID NO: 11 SEQ ID NO: 3
MGFFLFSOMPSEFLVSTLLFLIISHSHAQNSRMSLNTSGLGASTMOISIGGAGGNGLLGTSRONAGL
GGNSALGLGGGNQNTVNOGLAGLITCMMMMSMMGGGLMGGGLGGLGNGLGGGGLGEGLSNALNDML
CGSLNLGSKGGNNSTSTNSPDLQALGINTSQNDSTSGTSDTSDSDPQQQLLKMFEIMQSLFGDG
QDGTQSSSGGKOPTGEGQNAKKGVTDALSLGMLGSLQLLGGGGLGGGQGNAGTGLDGSLSGKGLQ
NLGGPDYQQLGNVAGTGTGMRAGTQALNDIGTHRSSTRSFVKNKGDRAKKEIGQFMQYPEVFGKPOY
QKPGQEVKTDKSWAKALSKPDDGDMTPASNEQFNKAKGMTKRPWAGDTGNGNLQARGAGGSSSLGIDAM
MAGDAINNHALGKLGAA1

Please check the joined
Sequences above to make sure
I got it right!
Thanks,
Thom

Point of Contact:
Thomas G. Larson, Ph.D.
703-308-7309
CM1, Rm. 6 B 01

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using 'sw model

Run on: January 22, 2003, 16:26:07 ; Search time 35 Seconds
(without alignments)
2572.645 Million cell updates/sec

Title: 693_11_3

Perfect score: 2248

Sequence: 1 MGFFLFQMPSEFLVSTLL.....DAMVAGDAINMALGKLGAA 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1181.5	52.6	382	2	Q9FCY8
2	1002.5	44.6	365	2	Q9KH45
3	840.5	37.4	356	2	Q93FY8
4	736.5	32.8	344	2	Q9EXP0
5	712	31.7	142	2	Q9LAW6
6	264.5	11.8	1489	16	O53559
7	255	11.3	591	16	O07224
8	254	11.3	622	16	O8VKN3
9	251.5	11.2	628	16	O8VJ19
10	250	11.1	635	16	O8VK71
11	250	11.1	775	16	O8VJ15
12	249.5	11.1	1217	16	O8VIY9
13	249	11.1	731	16	O50415
14	248	11.0	1381	16	O53552
15	248	11.0	1384	16	O8VI21
16	248	11.0	1715	16	O8VI20
					Q9FCY8 erwinia ste
					Q9KH45 pantoea agg
					Q93FY8 pectobacter
					Q9EXP0 erwinia chr
					Q9LAW6 erwinia amy
					O53559 mycobacteri
					O07224 mycobacteri
					O8VKN3 mycobacteri
					O8VJ19 mycobacteri
					O8VK71 mycobacteri
					O8VJ15 mycobacteri
					O8VIY9 mycobacteri
					O50415 mycobacteri
					O53552 mycobacteri
					O8VI21 mycobacteri
					O8VI20 mycobacteri

17	247.5	11.0	1079	16	O53557	mycobacteri
18	247	11.0	588	16	O50396	mycobacteri
19	246	10.9	1538	16	O53395	mycobacteri
20	245.5	10.9	923	16	O53890	mycobacteri
21	245.5	10.9	1665	16	O53215	mycobacteri
22	244	10.9	1507	16	O8VJ23	mycobacteri
23	241.5	10.7	879	16	O8VKD2	mycobacteri
24	241	10.7	3659	16	O98LN6	rhizobium l
25	238.5	10.6	882	16	O53845	mycobacteri
26	236.5	10.5	436	16	O8VIX6	mycobacteri
27	236.5	10.5	714	16	O53556	mycobacteri
28	236.5	10.5	853	16	O53439	mycobacteri
29	235.5	10.5	840	5	O95YF6	patinopecte
30	235	10.5	439	16	O06287	mycobacteri
31	234.5	10.4	1408	16	O8VK17	mycobacteri
32	234	10.4	2174	16	O92U08	rhizobium m
33	233.5	10.4	349	10	O9LSP2	arabidopsis
34	233.5	10.4	615	16	O05806	mycobacteri
35	232	10.3	837	16	O53684	mycobacteri
36	229.5	10.2	1329	16	O06810	mycobacteri
37	228.5	10.2	486	10	O9AR23	oryza sativ
38	228.5	10.2	667	16	O53415	mycobacteri
39	228.5	10.2	904	5	O76271	mytilus edu
40	228	10.1	783	16	O53809	mycobacteri
41	227	10.1	496	16	O53631	mycobacteri
42	227	10.1	533	16	O8VKR5	mycobacteri
43	226.5	10.1	491	16	O06818	mycobacteri
44	226.5	10.1	576	16	P71664	mycobacteri
45	226.5	10.1	650	16	O8VJW1	mycobacteri

ALIGNMENTS

RESULT 1

Q9FCY8 PRELIMINARY: PRT; 382 AA.
AC Q9FCY8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HrpN.
GN HrpN.
OS Erwinia stewartii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=66271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS104;
RA Frederick R.D., Ahmad M., Majerczak D.R., Arroyo-Rodriguez A.S.,
RA Manulis S., Coplin D.L.;
RT "Genetic organization of the Pantoea stewartii subsp. stewartii hrp
RT gene cluster and sequence analysis of the hrpA, hrpC, hrpN and wlsE
RT operons.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF282857; AAG01466.1; ...
SQ SEQUENCE 382 AA; 37878 MW; 8BA8C723F2DA38BB CRC64;

Query Match 52.6%; Score 1181.5; DB 2; Length 382;
Best Local Similarity 62.1%; Pred. No. 1.2e-60;
Matches 251; Conservative 44; Mismatches 86; Indels 23; Gaps 11;
QY 35 MSNTSLGCASTMQISIGGAGGNGNLGTSRQNAAGLGSALGLG-GGNQNDVTYNQLAGL 93
Db 1 MSNTSLGTSALQVTL---GGNGLMGTLDTDLGLGLSQPGLGEGKGHNESIDLLAAA 57
QY 94 LTGMAMMMVMGGLMGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGL 153
Db 58 LTGMAMMMVMGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGL 109
QY 154 NTSTSTNPLDQALGNTSTQNDDSTSGTSTSDSSDPMQOLLKMFSEIMQSLFGDGDG 213

```
Db 110 -TTGAGSS-----LGLDPTQTGDDSLSGAGQTSMS-PMEOIMKIPADITOSLFGD-QDG 161
QY 214 TCGSSSGKQPTGEQNAKKYKGVTDALSLGMNGLSQLLNGSLGGGCGGAGTGLDQSS 273
Db 162 ASGGNA-CRQPSODEONAYKKVTDALFAPMGGLSOVAGNSEGGLDGG-WGLG-CGNG 218
QY 274 LGGKGLQLNSPVYQOGLNAVGTGIGMKAGIOALNDICTHRSSTRSPVFNKGDRAK 333
Db 219 LGGKGLQLNSPADPQQLGNAICTGVGMKAGIEALNIGTHSDSSTRKSLINKEDRALARE 278
QY 334 IGFQMDQYPEVFGKPOYKQPGOEKVTDDKSWAKALSPPDDGMPASMPQNAKGMK 393
Db 279 VGFQMDQYPEVFGKPOYKQPNADSAVKIDTKSWAEALUSPDDGMPASMEQFNKAGIIX 338
QY 394 RPMAGDTGNGINLQARGAGSSGIGIDAMACDAINNMAKGLGAA 437
Db 339 SAMAGDNGINLQARGAGSSGICATLTCDAINNMAIRPLGAA 382

RESULT 2
ID Q9KH45 PRELIMINARY; PRT: 365 AA.
AC Q9KH45;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Harpin.
GN HRPN.
OS Pantoea agglomerans pv. gypsophilae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=48984;
RN 1;
RP MEDLINE=21171042; PubMed=11277443;
RA Mor H., Manulis S., Zuck M., Nizan R., Coplin D.L., Barash I.;
RT "Genetic Organization of the hrp gene cluster and dspAE/BF operon in
RL Erwinia herbicola pv. gypsophilae.";
RL Mol. Plant Microbe Interact. 14:431-436(2001).
DR EMBL: AF271716; AAF76342.1; -.
SQ SEQUENCE 365 AA; 36784 MW; 6F79B66DDDD95AC2A CRC64;

Query Match 44.6%; Score 1002.5; DB 2; Length 365;
Best Local Similarity 57.9%; Pred. NO. 2.3e-50;
Matches 223; Conservative 40; Mismatches 83; Indels 39; Gaps 14;

QY 35 MSLNTSGLCASTMQISIGAGGNNGLLTGTSRONAGLGNSALGGLGGNNDTVNOLAGLL 94
Db 1 MSLNTTPLGIPAMQISL-----GDSNGLPGPNLQNAVLN-----SMGDRSTDETINQLAGAL 52
QY 95 TGMNMMMSMMGGGLGGLGGGLGGLG--SCGLGCLSLNALNDMLCGSLNT-LGSKG 151
Db 53 TAMMMNMMMSGSL-GGGLGTA-GGELGGGAGGALGFLYNGOGLGGLGALGSLGGTG 110
QY 152 GNNTTSTTNSPLDQALGINSNDSTSGTSTSDSS-----DPMQQLKMFSEIMQSL 206
Db 111 G-----AGSILDRALDL-----DPTADGCLSDSNCTYCMSPIDQLMKMFAEVMQSM 157
QY 207 FGDGODGTQGSSESGKQPTGEQNAKKYKGVTDALSLGMNGLSQLLNGSLGGGCGGAG 266
Db 158 FG-GGGDTSGRGT-CNQPTPDBONAVTKGVTDALFAPMGGLSQQGS-GTGGHNGSIG 214
QY 267 TGLDSSLGKGLQLNSPVYQOGLNAVGTGIGMKAGIOALNDICTHRSSTRSPVFNK 326
Db 215 L---GNGLGGKGLGELNPADEQLGVGVGTGVMKAGIEALNIGTHSDSSTRFINKD 271
QY 327 DRAMAKEIGQFMDQYPEVFGKPOYKQPGOEKVTDDKSWAKALSPPDDGMPASMEQFN 386
Db 272 DRSIAKEVGQFMDQYPEIFGAPQYKTKTPIYSDVKTDTKSWAEALSNPDDGMPASMEQFN 331
QY 387 KAKGMIRPMAGDTGNGINLQARGAG 411
Db 332 KAKGMIKSAMSGDT--GNLQATGEG 354
```

RESULT 3

```
Q93FY8 PRELIMINARY; PRT: 356 AA.
AC Q93FY8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE HrpN.
GN HRPN.
OS Pectobacterium carotovorum subsp. carotovorum.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=555;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=SCC1.
RA Rantakari A.V., Virtaharju O., Vaehaeimiko S., Taira S., Palva T.E.,
RT "Type III secretion contributes to the pathogenesis of soft-rotting
RT plant pathogen Erwinia carotovora: Partial characterization of the hrp
RT gene cluster.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF302656; AAK97283.1; -.
SQ SEQUENCE 356 AA; 35534 MW; 154C9D885810DCB8 CRC64;

Query Match 37.4%; Score 840.5; DB 2; Length 356;
Best Local Similarity 48.0%; Pred. NO. 4.6e-41;
Matches 198; Conservative 36; Mismatches 114; Indels 59; Gaps 11;

QY 37 LNTSGLCASTMQISIGAGGNNGLLTGTSRONAGLGNSALGGLGGNNDTVNOLAGLLT 95
Db 2 LNSLGGAS-LQITI-KAGCNGGLFPSSQSGSPSQSAF---GGVRSNLAELQSLDIT 56
QY 96 GMMNMMMSMMGGGLGGLGGLG-----GGLGNGLGGSGGLGSLNALNDMLGGLSLTGS 149
Db 57 TMFMFMMSMMGGG--MSGGLGGLGSSLGGGLGGLG--GGLGGLGSSLGGSLGSALG--GG 111
QY 150 KGCNNTTSTTNSPLDQALGINSNDSTSGTSTSDSSDPMQQLKMFSEIMQSLFGD 209
Db 112 LGG-----ALGAGMAMNPMSMMGS-----LLFSALEDLLGG 143
QY 210 GQDGTGSSSGSGKQPTGEQNAKKYKGVTDALSLGMNGLSQLLNGSLGGGCGGAGTGL 269
Db 144 GMSQOQGLPENKQSPSPETSAVTQGVNDALSAILENGLSQTKGQ-----T 189
QY 270 DGSLSLGGKGLQLNSPVYQOGLNAVGTGIGMKAGIOALNDICTHRSSTRSPVFNKGDRA 329
Db 190 SPLQNGNGLGSLGAGAFNLGSLTGLMSYQKAGLQELANNISTHNDSPTRYFVDEKDRG 249
QY 330 MAKEIGQFMDQYPEVFGKPOYKQPGOEKVTDDKSWAKALSPPDDGMPASMEQFNKAK 389
Db 250 MAKEIGQFMDQYPEVFGKPEYKQDNMTAKODDKSWAKALSPPDDGMPGMSMDKPMKAV 309
QY 390 GMTKRPMDGDTGNGINLQARGAGSSGIGIDAMACDAINNMAKGLGA 436
Db 310 GMTAVAGDTGNTNLSARGNGCASLIDAMTGDRIVNMGLKLSS 356

RESULT 4
Q9EXP0 PRELIMINARY; PRT: 344 AA.
AC Q9EXP0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Harpin protein.
GN HRPN.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
```



```
Db 474 -----GGVGGVGGGAGAGAGHGGDG-----GSVNTPIGGSEAGDGK 515
Qy 403 GNQARGAGSSSLGIDAMMAGDAINNMALKLGAA 437
Db 516 GGLGGGGGRGIFG-QFGAGGAGGAGGVGGAGGA 548

RESULT 9
Q8VJ19
ID AC Q8VJ19 PRELIMINARY; PRT; 628 AA.
AC Q8VJ19;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PE_PGRS family protein.
GN MT3476.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007154; AAK47814.1; -.
DR TIGR; MT3476; -.
DR InterPro; IPR00104; Antifreeze_1.
DR InterPro; IPR001343; Hemlysn_Ca_bind.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR00308; ANTIFREEZEI.
DR ProDom; PD001223; PE_region; 1.
DR PROSITE; PS00583; PFKB_KINASES; 1; UNKNOWN_1.
SQ SEQUENCE 628 AA; 53828 MW; 9F62B1E7CC2C567E CRC64;

Query Match 11.28; Score 251.5; DB 16; Length 628;
Best Local Similarity 30.68; Pred. No. 6.5e-07;
Matches 118; Conservative 16; Mismatches 155; Indels 97; Gaps 21;

Qy 52 GGAGG--NNGLLGTSRQNALGGNSALGLGGNQNDTVNQLAGLLTGMMMSMMGGGGL 109
Db 320 GGAGGDANGNPANTSIANAGAGNGAAG-GDGGANGAGGAGGQAASGVGGDGGNGG 378
Qy 110 MGG----GLGGGLNGLGGSGGLGGL-SNALNDMLGSLNTLGSKGNNTTSTNSPLD 164
Db 379 AGGTGTGTHAGGAG-GAGGAGGGRGLVYNGNGNGCN----GAAGNGAIGGTGG--- 430
Qy 165 QALGINTSQNDSTGTDSTSDSDPQQLLKMFSEIMQSLFGDGDGTQGTSSSGGKQP 224
Db 431 -AGGV-PANQGNNSALGT-----QPVGDDGDDGNGGTGG----- 463
Qy 225 TEGQNAYKKVTDALSG-LMGNGLSOLLGNGGLG--GGQGNACTGLDGLSGKGLQN 281
Db 464 TGGRGDGGSGAGAGSLWMNGNG--GNGGTGGSGVGVNGGIGDGGAG-GGNATST 520
Qy 282 LSGPVDYQOLGNAYGTGTMKAGIQAINDIGTHRSSTRSFVNKGDRAKMAKEIGQFMQY 341
Db 521 SSIPFDAHG-GNG---GAGGDAG-----HGGTGGDGGDGGHAGTGGRGGL--- 562
Qy 342 PEVFGKPOYQKPGQEVKTDKSNAKLSKPDGDMTPASMEQFNKAKGMTKRPWAGDTG 401
Db 563 -----AGQHANSNGGGGGGTGGAGTGHG-TPGSGN-----AGGTG 596
```

```
Qy 402 NGNLQARGAGSSSLGIDAMMAGDAIN 427
Db 597 TGNADSTNGPGSDG-----LGGDAFN 618

RESULT 10
Q8VK71
ID Q8VK71 PRELIMINARY; PRT; 635 AA.
AC Q8VK71;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PE_PGRS family protein.
GN MT1123.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE005992; AAK45381.1; -.
DR TIGR; MT1123; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
SQ SEQUENCE 635 AA; 51726 MW; A461A61100C7C5D4 CRC64;

Query Match 11.18; Score 250; DB 16; Length 635;
Best Local Similarity 27.28; Pred. No. 8e-07;
Matches 120; Conservative 23; Mismatches 152; Indels 146; Gaps 20;

Qy 41 GLGASTMQISIGGAGNGNLLGT---SRQNALGGNSALGLGGNQNDTVNQLAGLLTGM 97
Db 293 GSGGHALLMGAGGAGGNGSGGTGGAGGTAGAGNGGAGGGGGT-----GGLLFG- 343
Qy 98 MMSMMMGGLMG-----GGL-----GGGLG--NGLGSGGLGEGLSNALNDMLG 141
Db 344 -----NGAGGGGATAGAGGAGANGVSTNGGTGGNGGIGGTGGGGAGGNA--GLLG 395
Qy 142 -----GSLNTLGSKGNNTTSTNSPLDQALGINTSQNDSTGTDSTSDSDPQQ 194
Db 396 VGGAGGHASGGADRGAGGTGFISS--DGGAGDGGDGGNGGAGGTGGL----- 444
Qy 195 LKMFSEIMQSLFGDGDGTQGTSSSGGKQPTEGEQNAYKKVTDALSLGMNG----- 247
Db 445 -----LPGAGNGGPG-GSGGAADIGNGGAGNGGTDGNGCNGSGGGAGSGG 492
Qy 248 -----LSOLLNGL--GGQGN-AGTGLDGLSGKGLQNLGSPVDYQOLGNAYG 296
Db 493 DGGAGGNGAWLFGNGGAGGGGKGGNGAGGLGGSGFLPGL-NGSG-----GDGD 544
Qy 297 TGIKMGAGIQAINDIGTHRSSTRSFVNKGDRAKMAKEIGQFMQYPEVFGKPOYKPGQ 356
Db 545 GGNAPGGVLYNG-----GAGGQSSGGIG-----GPG- 573
Qy 357 EVKTDKSNAKLSKPDGDMTPASMEQFNKAKGMTKRPWAGDTGNGNLQARGAGSSSLG 416
Db 574 -----ATGAGGKGGDGG-----DAQLIGDGGNGGNGAGGTGGTGP 610
Qy 417 IDAMMAGDAINNMALKLGAA 437
Db 611 PGPGGGGGLGGLLFGGTGTA 631
```


[illegible]

Search completed: January 22, 2003, 16:27:39
Job time : 39 secs

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	2050	91.2	403	1	HRPN_ERWAM	Q01099	erwinia amy
2	828.5	36.9	356	1	HRPN_ERWCA	Q47279	erwinia car
3	718.5	32.0	340	1	HRPN_ERWCH	Q47278	erwinia chr
4	248	11.0	463	1	YA68_MYCTU	O53416	mycobacteri
5	242	10.8	778	1	YQ34_MYCTU	P71933	mycobacteri
6	241.5	10.7	957	1	Y278_MYCTU	P56877	mycobacteri
7	238.5	10.6	515	1	Y140_MYCTU	Q50594	mycobacteri
8	238	10.6	1901	1	Y208_MYCTU	O53553	mycobacteri
9	235.5	10.5	914	1	WA22_MYCTU	O06794	mycobacteri
10	234.5	10.4	543	1	Y931_MYCTU	O50630	mycobacteri
11	234.5	10.4	603	1	YD25_MYCTU	Q10637	mycobacteri
12	233.5	10.4	338	1	GRP_ARATH	P27483	arabidopsis
13	232	10.3	801	1	Y747_MYCTU	O53810	mycobacteri
14	229	10.2	491	1	YK98_MYCTU	Q10707	mycobacteri
15	217	9.5	384	1	GRP1_PETHV	P09789	petunia hyb
16	213.5	9.5	481	1	LORI_MOUSE	P18165	mus musculus
17	213	9.5	1156	1	GLH4_CAEEL	O76743	caenorhabdi
18	211.5	9.4	747	1	SPD1_NEPCL	P19837	nephila cla
19	202.5	9.0	498	1	Y118_MYCTU	Q50615	mycobacteri
20	202	9.0	937	1	HYR1_CANAL	P46591	candida alb
21	197.5	8.8	344	1	POPA_RALSO	Q9rbs0	ralstonia s
22	196.5	8.7	465	1	GRP2_PHAVU	P10496	phaseolus v
23	193	8.6	183	1	GRP2_ORISA	P29834	oryza sativ
24	191	8.5	252	1	GRP1_PHAVU	P10495	phaseolus v
25	191	8.5	316	1	GRP1_HUMAN	P23490	homo sapien
26	190	8.5	641	1	EBN1_EBV	P30211	epstein-bar
27	187.5	8.3	1387	1	TROP_HUMAN	Q12816	homo sapien
28	187	8.3	959	1	N100_YEAST	Q02629	saccharomyc
29	186.5	8.3	1113	1	N116_YEAST	Q02630	saccharomyc
30	184.5	8.2	419	1	CSP_PLACM	P08676	plasmodium
31	177.5	7.9	1224	1	PER_DROME	P07663	drosophila
32	175.5	7.8	734	1	YKR2_CAEEL	P34308	caenorhabdi
33	174.5	7.8	672	1	PHX5_MOUSE	P09399	mus musculus

```

Db 61 TGMWMMWMMGGGLNGGLGGGLGNGGLGGGLGEGLSNALNDMLGGSINTLGSKGNN 120
QY 155 TTSTTNSPLDQALGINSTSONDDSTGCTSDTSDSPMQQLKMFSEIMOSLFQDQDGT 214
Db 121 TTSTTNSPLDQALGINSTSONDDSTGCTSDTSDSPMQQLKMFSEIMOSLFQDQDGT 180
QY 215 QGSSSGGKQPTGEQNAKGYVTDALSLGMLNGLSOLLGNGGLGGGGNAGTGLDSSSL 274
Db 181 QGSSSGGKQPTGEQNAKGYVTDALSLGMLNGLSOLLGNGGLGGGGNAGTGLDSSSL 240
QY 275 GKGGLNLSPVDYQOLGNVAGTGIGWKAGIOALNDIGTHRHSSSTRSFYKNGDRAMAKEI 334
Db 241 GKGGLNLSPVDYQOLGNVAGTGIGWKAGIOALNDIGTHRHSSSTRSFYKNGDRAMAKEI 300
QY 335 GOFMDQYPEVFGKPOYKQPGQEVKTDDKSWAKALSKPDDGTMTPASMEOFNKAQMIKR 394
Db 301 GOFMDQYPEVFGKPOYKQPGQEVKTDDKSWAKALSKPDDGTMTPASMEOFNKAQMIKS 360
QY 395 PHAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437
Db 361 AMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 403

```

RESULT 2

```

HRPN_ERWCA
ID HRPN_ERWCA STANDARD; PRT; 356 AA.
AC Q47279;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Harpin (Harpin-ECC) (Fragment).
GN HRPN.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=71;
RX MEDLINE=96405946; PubMed=8810071;
RA Cui Y., Medi L., Mukherjee A., Dumenyo C.K., Chatterjee A.K.;
RT "The RsmA mutants of Erwinia carotovora subsp. carotovora strain
RT Ec971 overexpress hrpNec and elicit a hypersensitive reaction-like
RT response in tobacco leaves.";
RL MOL. Plant Microbe Interact. 9:565-573(1996).
CC -!- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT
CC UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO
CC REQUIRED FOR PATHOGENICITY IN HOST PLANTS.
CC -!- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L78934; AAA09733.1; -
KW Hypersensitive response.
FT DOMAIN 1 224 GLY-RICH.
FT NON_TER 356 356
SQ SEQUENCE 356 AA; 35621 MW; 108B46B9D27F9DE4 CRC64;

```

```

Query Match 36.9%; Score 828.5; DB 1; Length 356;
Best Local Similarity 48.2%; Pred. NO. 1.5e-43;
Matches 196; Conservative 37; Mismatches 115; Indels 59; Gaps 11;

QY 37 LNTSLGLASTMISTGAGGNNGLL-GTSRQNRAGLGGNAGLGGGNQNDTYNQLAGLLT 95
Db 2 LNSLGGGAS-LOITI-KAGNGGLFPSSQNGSPSQSAF---GQGRNIAPQLSDIMT 56
QY 96 GMMWMMWMMGGGLNGGLG-----GGLGNGLGGGGLGEGLSNALNDMLGGSINTLGS 149

```

```

Db 57 TMMPGCMGMMGG--MSGCLGCLGSLGGLGGGLG-GGLGGGLGSSGSGGSLG--GG 111
QY 150 KGGNNTTSTTNSPLDQALGINSTSONDDSTGCTSDTSDSPMQQLKMFSEIMOSLFQD 209
Db 112 LGG-----ALGAGMNAFSAKMS-----LFSALEDLGG 143
QY 210 GQDTGSSSGGKQPTGEQNAKGYVTDALSLGMLNGLSOLLGNGGLGGGGNAGTGL 269
Db 144 GMSOQGGGLFKNQFSSPEISAYTQGVNDALSAILGNGLSOTKGO-----T 189
QY 270 DGSLSGKGLQNLSPVDYQOLGNVAGTGIGWKAGIOALNDIGTHRHSSSTRSFYKNGDRA 329
Db 190 SPLQNLGNGLOGLSGAGAFNQLGSLGMSVGGKAGLOELNINISTHNSDPTFYVDKEDRG 249
QY 330 MAKEIGQFMQDQYPEVFGKPOYKQPGQEVKTDDKSWAKALSKPDDGTMTPASMEOFNKA 389
Db 250 MAKEIGQFMQDQYPEVFGKAEYKQDNWQTAQEDKSWAKALSKPDDGTMKGMKPFKAY 309
QY 390 GMTKRPMDGTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 436
Db 310 GMIKSAIRGDTGNTLSARGNGGASLGDRAAMIGDRIVNMGELKKLSS 356

```

RESULT 3

```

HRPN_ERWCH
ID HRPN_ERWCH STANDARD; PRT; 340 AA.
AC Q47278;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Harpin (Harpin-ECH).
GN HRPN.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC16 / AC1450;
RX MEDLINE=961172740; PubMed=8589405;
RA Bauer D.W., Wei Z.M., Beer S.V., Collmer A.;
RT "Erwinia chrysanthemi harpinEch: an elicitor of the hypersensitive
RT response that contributes to soft-rot pathogenesis.";
RL MOL. Plant Microbe Interact. 8:484-491(1995).
CC -!- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT
CC UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO
CC REQUIRED FOR PATHOGENICITY IN HOST PLANTS.
CC -!- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L39897; AAC31978.1; -
KW Hypersensitive response.
FT DOMAIN 1 203 GLY-RICH.
FT NON_TER 340 AA; 34274 MW; E4D82C23731EF4C5 CRC64;
SQ SEQUENCE 340 AA; 34274 MW; E4D82C23731EF4C5 CRC64;

```

```

Query Match 32.0%; Score 718.5; DB 1; Length 340;
Best Local Similarity 42.0%; Pred. NO. 6.4e-37;
Matches 172; Conservative 41; Mismatches 114; Indels 77; Gaps 10;

QY 47 MQLSI-GGAGNNGLLGTSRQNRAGLGG-NSA---LGLGGGNQNDTYNQLAGLLTGMWMM 101
Db 1 MQITIRAHIGDGLGVSGGLGCAQGLRGLNSAASLGSVDKLSSTIDKLTSLTSMW--- 57
QY 102 SMWGGGGLMGGLGGGLGNGLGGSGGLGEGLSNALNDMLGGSINTLGSKGNNTTSTNS 161

```

Db 58 -----FGALAGLGAS-SKGLGMSNLQCSFGN-----GAQASNLLSPVK- 98

Qy 162 PLDQALGINSTSDNSTSGTDTSDSDPMQOLKMFSEIMQSLFG-----DQDQG 213

Db 99 -----SGDALS-----KMFKALDLDLGHDTVTKLTQSNQ 130

Qy 214 TQSSSGGKQPTGEQONAYKKVTDALSLMGNLSQLLNGGLGGQGGNAGTGLDGG 273

Db 131 LANSMLNASQMTQGNNAFSGVNNALSSILGNLQGSW-----SGFSQPS 176

Qy 274 LGGKQLNLSPVDYQOLGNNAVGTGIGMKAGIQALNDIGCTHRHSSPSRFSVKNKGDRAK 333

Db 177 LGAGLGLSGAGAFNLGNALGMVGQNALSLNSVTHVDGNRHFVDKEDRAK 236

Qy 334 IGQFMDQPEVFGKPYQKQPGQYKTKDKSWAKALSKPDDDMTPASMEQFNKAKGM 393

Db 237 IGQFMDQPEVFGKPYQKQPGQYKTKDKSWAKALSKPDDDMTPASMEQFNKAKGM 296

Qy 394 RPMAGDTGNLQARGAGSSGIGDAMMAGDANNWALGKGA 437

Db 297 SAVAGDTGNTNLNRGAGGASLGIDAAYVVDKRIANNLSGLKLANA 340

RESULT 4

YA68_MYCTU STANDARD; PRT; 463 AA.

AC O53416;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical PE-PGRS family protein Rv1068c.

GN Rv1068C OR MT1097 OR MTV017.21c.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37Rv;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., McLean J., Moule S., Murphy L.,

RA Rutter S., Seeger K., Skelton S., Squares R., Rogers J.,

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RA "Deciphering the biology of Mycobacterium tuberculosis from the

RA complete genome sequence.";

RA Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RA "Whole genome comparison of Mycobacterium tuberculosis clinical and

RA laboratory strains.";

RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RL -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AL021897; CAAL17184.1; -.

DR EMBL; AE006991; AAK45353.1; ALT_INIT.

DR TIGR; MT1097; -.

DR Tuberculist; Rv1068c; -.

DR InterPro; IPR000084; PE_region.

DR Pfam; PF00934; PE; 1.

DR ProDom; PD001223; PE_region; 1.

KW Hypothetical protein; Complete proteome.

FT CONFLICT 218 218 T -> S (IN REF. 2).

FT CONFLICT 235 235 G -> GGGAGIGGADCTKGGDAGCAGGAGGWIHHGGV

FT GGGGTGGGGDGVQGPEDTGAAGGAGG (IN REF. 2).

FT

SQ SEQUENCE 463 AA; 39305 MW; CF5696A7E9593952 CRC64;

Query Match 11.0%; Score 248; DB 1; Length 463;

Best Local Similarity 31.8%; Pred. No. 2.4e-08;

Matches 93; Conservative 17; Mismatches 112; Indels 70; Gaps 13;

Qy 42 LGASTMQISIGGAGGNLLGTSTRONAGLGNLSALGGLGGGNQNDTVNQLAGILTCMMMM 101

Db 203 IGAPGVAGGAGGAGGTAGLFG-----NGCAGCAGGAGGAGGRGD--GGAGNLSONGDA 256

Qy 102 SMMGGGLMGGLGGGLGNLGGSGGLGEGLSNALNDMLGSLNTLGSKGNNTTSTNS 161

Db 257 GTGGGGNAGNGGNGGSAGWLSNGGTGG-----GCTAGA-GGQGGNG-----NS 301

Qy 162 PLDQALGINSTSDNSTSGTDTSDSDPMQOLKMFSEIMQSLFGDQDGTQSSS--- 218

Db 302 GIDPG---NGGQADTGNAGNGHGSAA-----AKLFGDGGAGGAGMGST 344

Qy 219 -----SGGKQPTGEQONAYKKVTDALSLGMLGNLSQLL-----GNGGLGGQG--G 263

Db 345 GTGGGGGFGGTGGNGNGHAGGAGSGGTAGLLGSGSGGTGGDGGNGLGAGSGAKG 404

Qy 264 NAGTGLDSSLGKGLQNLGSPVDYQOLGNNAVGTGIGMKAG---IQALNDIG 312

Db 405 NGNGGD-----GGKG-----GDAQLIGNGGNGGKGGTGLMPLGNGTG 445

RESULT 5

YQ34_MYCTU STANDARD; PRT; 778 AA.

ID YQ34_MYCTU

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE Hypothetical PE-PGRS family protein Rv2634c.

GN Rv2634C OR MT2712 OR MTCV441.04C.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37Rv;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., McLean J., Moule S., Murphy L.,

RA Rutter S., Seeger K., Skelton S., Squares R., Rogers J.,

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RA "Deciphering the biology of Mycobacterium tuberculosis from the

RA complete genome sequence.";

RA Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RA "Whole genome comparison of Mycobacterium tuberculosis clinical and

RA laboratory strains.";

RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RL -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

```

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PERS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z80235; CAB02341.1; -
CC EMBL; AE007103; AAK47026.1; ALT_INIT.
CC TIGR: MT0212; -
CC TubercuList; RV2634c; -
CC InterPro: IPR000084; PE_region.
CC Pfam: PF00934; PE; 1.
CC ProDom: PD001223; PE_region; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 51 51 V -> L (IN REF. 2).
FT CONFLICT 63 63 Q -> H (IN REF. 2).
FT CONFLICT 274 274 A -> T (IN REF. 2).
SQ SEQUENCE 778 AA: 63131 MW: DAB20FE58E4999E7 CRC64;
Query Match 10.8%; Score 242; DB 1; Length 778;
Best Local Similarity 30.1%; Pred. No. 9.8e-08;
Matches 91; Conservative 19; Mismatches 110; Indels 82; Gaps 14;
QY 41 GLGASTMQISIGAG--GNNLLGTQRNAGLGGNSALGGV-----GNQDNTVNLQAGLL 94
Db 353 GNGGNANWFSCGAGCGGTGLACTGNVPCSTANTNGANGTDSNGNQV-----404
QY 95 TGMMMMSMMGGGLGGGLG--NGLGGSGGLGEGLSNALNDMLGSLNLTGSKGG 152
Db 405 -----GGNG--GPGFAGVGGEAGVGGGGGLGESLDG--NDGTGG-----KGS 443
QY 153 NNTTSTNSPLDQALGINSTQNDSTSGTDSSTSDSPMQQLKMFSETMOSLFDGQD 212
Db 444 AGGTAGTGGAGGAGGAGGIGETDGSAGV-----ATGGEAGG 481
QY 213 GTGGSSGKQPTGE-QNAYKKGVTDAL---SGLMNG---LSQLLNGGLGG-----259
Db 482 GATGGVGVGGAGGKCGKCGCHNFGVGDATCGDGGICGDCGNCALCAAGGNGGTGGAGNGC 541
QY 260 -----GCGGNAGTGLDSSLGKGLQNLSPV---DYQQLGNVAVGTGIGMKAGIOALND 310
Db 542 RGMMLNGGAGGAGGTGCT-GGGGAAGFAGVGAGGEGELTDGAGTAEAGTGGGLGG 600
QY 311 IG 312
Db 601 VG 602
RESULT 6
Y278_MYCTU
ID Y278_MYCTU STANDARD; PRT; 957 AA.
AC P56877;
DT 30-MAY-2000 (Rel. 39, Created)
DI 30-MAY-2000 (Rel. 39, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PERS family protein RV0278c precursor.
GN RV0278C OR MT0291 OR MT035.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID-1773;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Bacheam D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seecher K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
KL Nature 393:537-544(1998).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy K., Dodson R., Gwinn M.D., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PERS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL021930; CAAL7353.1; -
CC EMBL; AB006936; AAK44511.1; ALT_INIT.
CC TIGR: MT0291; -
CC TubercuList; RV0278c; -
CC InterPro: IPR000084; PE_region.
CC Pfam: PF00934; PE; 1.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 957 HYPOTHEITICAL PE-PERS FAMILY PROTEIN
FT SIGNAL 40 40 M -> I (IN REF. 2).
FT CONFLICT 158 163 MISSING (IN REF. 2).
FT CONFLICT 807 807 R -> G (IN REF. 2).
SQ SEQUENCE 957 AA: 81905 MW: 71EBAD417FBA47C CRC64;
Query Match 10.7%; Score 241.5; DB 1; Length 957;
Best Local Similarity 25.8%; Pred. No. 1.3e-07;
Matches 117; Conservative 42; Mismatches 149; Indels 145; Gaps 21;
QY 36 SLATSGLGASTMQISIGAGGAGNGLLGTSRQNALGG-----NSALGLGGNQDNTVNLQ 91
Db 362 TLTAGGIGGA-----GGAGGNAGLLFGSGSGGAGGFGADGGCGGPGGN-----A 407
QY 92 GLLTGMMMSMMGGGLMG--GGLLGGGLNGLGSGGL-----GEGLSNAL---N 137
Db 408 G-----TVFGSGGAGGNGVGGVGGFAGGIGAGGTGGLNGNGNGNGGSAVTTGN 458
QY 138 DMLCGSLNTLCSKGCNNNTTSTNSPLDQALGINSTQNDSTSGTDSSTSDSPMQQLK 197
Db 459 GGIGGTGVLLIGN--GGNGSGGIGAGKAGAGVGVSGLLLGLDGFNAPASTSPHITLQNNVLN 517
QY 198 MFSEIMQS-----LFGDQDQCTQCS-----SSGKQKPTGEQKNAYKKGVTDALS--LMNG 247
Db 518 VVNEPFTQLGRPLIGNANGTPTCTGADGGAGGWLFGNGANGTPTCTGAAGGAGGLFGNG 577
QY 248 LSQLLNGGLGG-----GCGGNAGTGLDSSSLGKGLQNLUSGPVDY 288
Db 578 -----GNGGHGATNTAATATGAGGAGGAGGILFTGGNGGTG--GIATGAGGTGAGGAGGV 630
SEQUENCE FROM N.A.

```



```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AL022022; CAAL17745.1; -.  
DR Tuberculin; RV3508; -.  
DR InterPro; IPR000084; PE_region.  
DR Pfam; PF00934; PE; 1.  
DR ProDom; PD001223; PE_region; 1.  
KW Hypothetical protein; Repeat; Signal; Complete proteome.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN  
FT CHAIN RV3508.  
SQ SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;  
  
Query Match 10.6%; Score 238; DB 1; Length 1901;  
Best Local Similarity 27.6%; Pred. No. 4.4e-07;  
Matches 124; Conservative 29; Mismatches 180; Indels 116; Gaps 23;  
  
QY 43 CASTWMOISL-GGAGGNNGLGFSRONAGLGGNSALGLGGGNDVTVNOLAGLLTGMMMM 101  
DB 1133 GASTTSINAGGAGNGGTGG-----KGGAGGAGTGVGGSGGTGGDGDAG----- 1179  
QY 102 SMWGGGGLGGGLG--GGLGN-GLGSGGLG--EGLSNALNDMLGSLNPLGSKGQNNTPS 157  
DB 1180 --SGGGCGGCAACKAGCGGCGVGGGEGCAGCLGLCLSGFDGCG---GGGGAGGCSA 1233  
QY 158 TTNPLDQALGINSTSDSTSGTSDTSDSDPQQOQLLMFSEIMQSLFG--DGQDGTQ 215  
DB 1234 -----GAGGINGAG---GAGGTGGAGGCGAP-----ATLGGPDGDCGQ 1270  
QY 216 GSSSCGKOPTGEQNAVYKGVY----DALSLGLMGLSOLLNGLG--GGGGNAGTGL 269  
DB 1271 GGIGG-----DGGNAGFAGVPGDGGGAGFAGVPG---GDGGTGGTGGAGGAGAGA 1322  
QY 270 DG-----SSLGKGLQLNLSGPVDYQO-----LQNA 294  
DB 1323 DGDPSIDGGCGGAGGCGGCGKGLNSTGLASASDGGNGGAGGAGGNGDGDGFIGS 1382  
QY 295 VGT-GIGMKAGIQAL-NDIGTHRHSSSTRSFYKNGDRAMAKEIGQFMQDYFEYFGRPQYK 352  
DB 1383 GCTGTGSDAGVGLANTGCTAGNAGIGAGGCGGDDGAGSDGALSDQNGCFAGCGGCG 1442  
QY 353 GPCQEVKTDKSWAKALSKPDDDCMTPASMEQFNKAKMKRPM-----GDTGNGNLQAR 408  
DB 1443 GYGNAGAGGINGAGGTG-----GTGGAGGCGQNTTGVASEGGAGGCGGCGGCGGI--- 1494  
QY 409 GAGSSSLGIDAMACDAINNALSLGLAA 437  
DB 1495 GGAGNAGFGAGVPDGG-----GIGGTGGA 1519  
  
RESULT 9  
WA22_MYCTU  
ID WA22_MYCTU STANDARD; PRT; 914 AA.  
AC O06794;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE WA22 antigen precursor.  
GN WA22 OR RV1759C OR MT1807 OR MTCY28.25C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmler K., Gao S., Barry C.E. III, Tekla F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Heiroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
```

```
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RL complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS  
CC SUBFAMILY.  
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
CC FRAMESHIFT IN POSITION 85.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; Z95890; CAB09322.1; -.  
DR EMBL; AE007040; -; NOT_ANNOTATED_CDS.  
DR TIGR; MT1807; -.  
DR Tuberculin; RV1759C; -.  
DR InterPro; IPR000084; PE_region.  
DR Pfam; PF00934; PE; 1.  
KW Antigen; Repeat; Signal; Complete proteome.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 914 WA22 ANTIGEN.  
SQ SEQUENCE 914 AA; 74354 MW; F6953C3DBE8E6AC8 CRC64;  
  
Query Match 10.5%; Score 235.5; DB 1; Length 914;  
Best Local Similarity 25.0%; Pred. No. 2.9e-07;  
Matches 120; Conservative 38; Mismatches 173; Indels 135; Gaps 20;  
  
QY 41 GLGASTMQISIGGAGGNG-----LLGTSRONAGLGGNSAL-----GLGGNQNDTVNQ 89  
DB 186 GAGGSMLFGAGGAGGAGGAATSLVGGTGGTGGNAGMLAGAGGAGGFSFTAGC 245  
QY 90 LAGL-LTGMMMMMMGGGLMG-----GGLGGGLG--NGLGGSGGLGE----GLSNA 135  
DB 246 AGGAGGAGGLFTTGGVGGAGGQHTGGAGGAGGAGGLFGAGGNGAGGFGDHTLCTGGA 305  
QY 136 LNDMLGSLNPLGSKGQNNTPSTNPLDQALGINSTSDSTSDSDPQQOQL 195  
DB 306 GGDGGCGGGLFGAGGCGGCGSLTTGGA-AGNCGNAGTSLGAGGAGGTGGAGG----- 359  
QY 196 LKMFSLMQSLFGSDGQDGTGS-----SSGKQPTGEQNAVYKGVTDALSLMGNG 247  
DB 360 -----TVFGGGKGGAGCGACGNACMLFCGCGCGGTGG-----FGAAGCGGCGGGS 404  
QY 248 LSOLLNGLGGGSGG-----GNAGTGLDSSSLGGKGLQLNLSGP 285  
DB 405 AGMLSGSGSGGAGGSGGFPAGTAAGAGGAGGAGPGLIGNGNGNGGSGGCGTGVGAG- 463  
QY 286 VDYQQLGNVAGTGTGMKAGTQA-----LNDIGTHRHSSSTRSFVNKGDRA 331  
DB 464 -----GNVLIGNGCGEGGIGALAGKSGFGGFGGLLLGADYNAPESTSPWHN-----LQ 512  
QY 332 KEIQGMDQYPEVF-GKPYQKGFQGVETDDKSWAKALSKPDDDCMTPASMEQFNKAK 390  
DB 513 QDTLSFINEPTALTRPLTQNG-----DSGTPQTGDD-CCAGC 550
```



```
CC      Or send an email to license@isb-sib.ch).
CR      -----
DR      EMBL; 273902; CAA98089.1; -.
DR      EMBL; AB007010; AAK45630.1; ALT_INIT.
DR      TIGR; MT1367; -.
DR      TubercuList; Rv1225c; -.
DR      InterPro; IPR000084; PE_region.
DR      Pfam; PF00934; PE; 1.
DR      ProDom; PD001223; PE_region; 1.
DR      *KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT      SIGNAL          1   30
FT      CHAIN           31   603
FT                                     HYPOTHETICAL PE-PCRS FAMILY PROTEIN
FT                                     Rv1325c.
FT      DOMAIN          114   603
FT      CONFLICT        132   132      G -> V (IN REF. 2).
FT      CONFLICT        135   135      G -> D (IN REF. 2).
FT      CONFLICT        337   337      G -> D (IN REF. 2).
FT      CONFLICT        508   508      L -> F (IN REF. 2).
FT      CONFLICT        508   508      L -> F (IN REF. 2).
SQ      SEQUENCE        603 AA; 49575 MW; 4F9BCB82B07AE964 CRC64;

Query Match          10.4%; Score 234.5; DB 1; Length 603;
Best Local Similarity 25.9%; Pred. No. 2.1e-07;
Matches 116; Conservative 35; Mismatches 142; Indels 155; Gaps 24;

QY 20 LFLIISHSHSQNSRMSLN-----TSLGASTPMQISIGGACGNN--GLLGTSRQNAAGLG 72
Db 247 LFEVLAAGGAGGTGLSVNGTGGTGGTGGGLFSSNGGAGGAGGFGVSGSAGGNGTGG 306
QY 73 NSAL-----GLGGGNQNTVQLAGLLTGMMMMMMGGGG-----LMG-CGLGQ-----G 117
Db 307 DGGTPTNGCGTCTGCTGCTGNQLVGGEGG-----ACGACGNAGTLFCAGGIGTGGTG 359
QY 118 LG-----NLGGGGLGGLGSLNALNDMLGSLNLTGSKGNNNTTSTTNSPLDQALGINS 173
Db 360 LGAPPDGTGKGGVG-GIGCA-----GALPGCGAGGTG-----GFGASSAD 401
QY 174 QNDSTSTGTDSTSDSPMQQLKMFSEIMOSLFGDGDGTGSSSGKOPTGEONAYK 233
Db 402 QMAGGIGSGSGGAA-----KLIGDGGAG-----GTGG-----430
QY 234 KGVTDALSGLNGLSQLLNGGLG-----GGGAGNAGTCLDSSLGKGLONLSQPDVY 288
Db 431 ----DSVRGAAGSG-----GTGGTGLLDGAGGAGGTGIEFGSVGGAGGAG-----474
QY 289 QQLGNNAV-GTIGMKAGTQALNDICTHRHSSTSFVNKGDRAKAIQFMQYPEVFGK 347
Db 475 ---GNAAGLGGAGGAGGAGGEE-----TAGDCAAGNAGLL-----508
QY 348 POYKPGQGEVKTDDKSWAKALSKPDDGDMTPASMEQFNKAKGMIRKPMAGDTGNGNLQA 407
Db 509 ----NGDG-----GAGGAGGLGIAGDGG-----NGGRKG-KAGMVGNGDGG-----544
QY 408 RCAGGSSLCIDAMMAGDAINNMALKLG 435
Db 545 -GAGGASVAVGNGVGGSGGNATLIGNG 571

RESULT 12
GRP_ARATH
ID GRP_ARATH STANDARD; PRT; 338 AA.
AC P27483;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Glycine-rich cell wall structural protein precursor.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OV. Columbia.
```

```
RX MEDLINE=92003708; PubMed=1912511;
RA Quigley F., Villiot M.L., Mache R.;
RT "Nucleotide sequence and expression of a novel glycine-rich protein
RL gene from Arabidopsis thaliana.";
RL Plant Mol. Biol. 17:949-952(1991).
CC -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X58338; CAA41249.1; -.
DR FIR; S17732; KNU0.
DR Cell wall; Structural protein; Repeat; Signal.
KW SIGNAL          1   20      POTENTIAL.
FT CHAIN           21   338      GLYCINE-RICH CELL WALL STRUCTURAL
FT                                     PROTEIN.
FT      DOMAIN          21   338      GLY-RICH.
FT      SEQUENCE        338 AA; 23891 MW; 046A6E8C1A4E89EB CRC64;
SQ

Query Match          10.4%; Score 233.5; DB 1; Length 338;
Best Local Similarity 29.9%; Pred. No. 1.3e-07;
Matches 80; Conservative 14; Mismatches 77; Indels 97; Gaps 10;

QY 41 GLGASTPMQISIGACGNNLLGTSRONA--GLGCNSALGLGGGNNONTVNLGAGLUTGMM 98
Db 59 GLG-----GGAGGGGIGGGAGGAGGGGLGGAGGGGLGGH-----GGGIGG-- 100
QY 99 MMSMGGGGLMGGGLGGGLGNGLGG--SGLGEGLSNALNDMLGSLNLTGSKGNNNTT 156
Db 101 -----GAGGAGGGGLGGHGGGGGGGGGGGLGGGTGGGAGGGGGGLGGH-- 152
QY 157 STTNSPLDQALGINSQNDSTSGTDSSTSDSPMQQLKMFSEIMQSLFGDGDGTG 216
Db 153 -----GGGIGGAG-- 161
QY 217 SSSGGKOPTGEONAYKGVTDALSLGMNGLSOLLGNGLGGGGGOGNAGTGLDSSSLGG 276
Db 162 GGAGG-----GLGGHGGGIGGGAGGSGGGGL-----GGGIGGAGGGAGCG--GGAGGG 209
QY 277 KGLQNLGSPVDYQQLGNNAVGTGICMKAG 304
Db 210 GGLGCGHG-----GCGFGGAGGGGLGGAG 233

RESULT 13
Y747_MYCTU
ID Y747_MYCTU STANDARD; PRT; 801 AA.
AC O53810;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PCRS family protein Rv0747 precursor.
GN Rv0747 OR MT0772.5 OR MY041.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
```



```
FT CONFLICT 312 312 C -> GC (IN REF. 1).
SQ SEQUENCE 491 AA; 41979 MW; 12C8630C59CA0C13 CRC64;

Query Match
Best Local Similarity 29.2%; Score 229; DB 1; Length 491;
Matches 106; Conservative 22; Mismatches 149; Indels 86; Gaps 17;

QY 43 GASTWQISIGAGGNNLLGTSRQAGLGNSALGLGGGNQNDTVNQLAGLL-----94
-bb 145 GDSFSPGVAGGAGSAGLIG-----NGRRGNGAPGGAGCN-----GGLGGLLLNGGGAGG 195
QY 95 --TGMNMMMSMMGGGLMGGGLGG-----GLGNGLGSGSGGLG 129
Db 196 VGTGDNVGLGAGGGGGGLGAGLHGAGAGNGCGCHGCGKAGSGSGSGGRC 255
QY 130 ECLSNALNDMLGGSILNLCGKGGNNNTTSTNSPLDQALGINSTSDST 179
Db 256 Q--FGGAGGLLYNGGAAGS--GGNGGDAGTGVSSDGFAGLGSGGSGGAGLIGVGGGGG 312
QY 187 DSSDP--MQQLLKMFE-----IMOSLPEDGDDCTQGSSSGCKQPTCEQONAYKCVTDA 239
Db 313 NCGDPLGLCARLFQVCSRGDGGVGGWLYGDGGGGDGGNGG--LPIGSTNA-----362
QY 240 LSLGMLNGLS--OLLGNGLGSGGGGNNAGTGLDSSLGKGLQNLGSPVDYQQLGNAYVTG 298
Db 363 ----GNGGSARLIGNGAGGSGSGGAPGCSVSSGCGVAGNPGGCGNGGVWYQNG--GAG 416
QY 299 ICMKAGIOALNDI-----CPIRHSSRFSVFNKCDRAKEIGOFMDQYPEVFGKPYQK 352
Db 417 GAAGGGGPGMNTTSPFGPGGPGVGGHGGTALFDGGAGGAGAGG-----P-----GTPDGAA 468
QY 353 GPG 355
Db 469 GPG 471

RESULT 15
GRP1_PETHY STANDARD; FRG; 384 AA.
AC P09789;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Glycine-rich cell wall structural protein 1 precursor.
GN GRP-1.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
RN [1]
RP SEQUENCE FROM N.A.
RA Condit C.M., Meagher R.B.;
RT "A gene encoding a novel glycine-rich structural protein of petunia.";
RL Nature 323:178-181(1986).
CC -!- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Cell wall (Potential).
CC -!- MISCELLANEOUS: THIS PROTEIN CONTAINS 67% GLYCINE RESIDUES.
CC -!- MISCELLANEOUS: 90% OF THE MATURE PROTEIN RESIDUES ARE CAPABLE OF
CC FORMING A BETA-PLEATED SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.
CC -!- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED
CC FAMILIES OF REPEATS, P1 AND P2, EACH REPEAT CONTAINING ABOUT 40
CC AA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04335; CAA27866.1; -.
```

```
DR PIR; A26099; A26099.
KW Cell wall; Structural protein; Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 384 GLYCINE-RICH CELL WALL STRUCTURAL
FT PROTEIN 1.
FT DOMAIN 41 384 GLY-RICH.
SQ SEQUENCE 384 AA; 28777 MW; C8541C549417D18C CRC64;

Query Match
Best Local Similarity 27.5%; Score 217; DB 1; Length 384;
Matches 86; Conservative 18; Mismatches 129; Indels 80; Gaps 11;

QY 1 MGFPLSQMPSFPLVLTLLFLIISHSHRAQNRMSLNTGCLCASTWQISIGAGGNNCL 60
Db 9 ICLLFFSSI--FFELTAITL-----ADKLEESRWG--NDNGCGFRRRGCGGGRGGRGPS 60
QY 61 LGTSR--ONAGLGNSALGLGGGNQNDTVNQLAGLLTGMMNMMMSMMGGGLMGGGLGGLG 119
Db 61 FGRGRAGAGGFGGAGGAGGG-----LGGGGGLGGGAGGG 98
QY 120 NGLGSGGLGCEGLSNALNDMLGSLNLTGSKGNNTTSTNSPLDQALGINSTSDST 179
Db 99 GGLGGGGAGGFGGAGGAGGAGGGLGGGGLGGGGGGG-----AGCGGVCGGAGSG 150
QY 180 SGTDSSTSDSPMQQLLKMFEIMQSLFCDDQDCTQGSSSGCKQPTCEQONAYKCVTDA 239
Db 151 GG-----FCAGGCGVGGGAGAGG---GVGGGGGFGGGGGG 182
QY 240 LSLGMLNGLSQLLNGSLGGGQGGNAGTGLDSSLGKGLQNLGSPVDYQQLGNAYVTGI 299
Db 183 VGGGSGHG-----GPGAGCGVGGGAGGGL--CGCVGGGGGGGSGG-----CGGIGGGS 299
QY 300 GMKAGIQALNDIG 312
Db 230 GHGGGFGAGGGVG 242

Search completed: January 22, 2003, 16:26:56
Job time : 17 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:26:07 ; Search time 21 Seconds
(without alignments)
2000.512 Million cell updates/sec

Title: 693_11_3
Perfect score: 2248
Sequence: 1 MGFFLFQMPFVSTLLI.....DAMMAGDAINNALGKLGAA 437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2050	91.2	403	2 T08471	harpin - Erwinia a
2	264.5	11.8	1489	2 D70807	hypothetical glyci
3	255	11.3	591	2 B70523	hypothetical glyci
4	249	11.1	731	2 C70974	hypothetical glyci
5	248	11.0	463	2 B70893	hypothetical glyci
6	248	11.0	1381	2 E70806	hypothetical glyci
7	247.5	11.0	1079	2 B70807	hypothetical glyci
8	247	11.0	588	2 F70971	hypothetical glyci
9	246	10.9	1538	2 B70846	hypothetical glyci
10	245.5	10.9	923	2 E70820	hypothetical glyci
11	245.5	10.9	1660	2 A70869	hypothetical glyci
12	242	10.8	778	2 F70963	hypothetical glyci
13	241.5	10.7	957	2 D70835	hypothetical glyci
14	238.5	10.6	515	2 B70663	hypothetical glyci
15	238.5	10.6	882	2 B70812	hypothetical glyci
16	238	10.6	1901	2 F70806	hypothetical glyci
17	236.5	10.5	714	2 A70807	hypothetical glyci
18	236.5	10.5	853	2 A70896	hypothetical glyci
19	235.5	10.5	914	2 D70987	hypothetical glyci
20	235	10.5	439	2 D70954	hypothetical glyci
21	234.5	10.4	543	2 F70726	hypothetical glyci
22	234.5	10.4	603	2 A70770	hypothetical glyci
23	234	10.4	2174	2 E95965	hypothetical glyci
24	233.5	10.4	338	1 KNUU	glycine-rich cell
25	233.5	10.4	434	2 E70768	hypothetical glyci
26	233.5	10.4	615	2 H70589	hypothetical glyci
27	232	10.3	801	2 F70824	hypothetical glyci
28	232	10.3	837	2 E70835	hypothetical glyci
29	229.5	10.2	1329	2 E70917	hypothetical glyci

30	228.5	10.2	667	2 A70893	hypothetical glyci
31	228	10.1	783	2 E70824	hypothetical glyci
32	227	10.1	496	2 H70839	hypothetical glyci
33	226.5	10.1	491	2 D70916	hypothetical glyci
34	226.5	10.1	576	2 A70900	hypothetical glyci
35	226	10.1	694	2 F70868	hypothetical glyci
36	225.5	10.0	1011	2 F70620	hypothetical glyci
37	225	10.0	749	2 A70812	hypothetical glyci
38	224	10.0	1306	2 A70934	hypothetical glyci
39	223	9.9	562	2 B70953	hypothetical glyci
40	222	9.9	396	2 T49109	glycine-rich prote
41	222	9.9	767	2 E70895	hypothetical glyci
42	221.5	9.9	639	2 D70931	hypothetical glyci
43	220.5	9.8	484	2 F70846	hypothetical glyci
44	220	9.8	532	2 F70580	hypothetical glyci
45	219.5	9.8	741	2 G70917	hypothetical glyci

ALIGNMENTS

RESULT 1

T08471
harpin - Erwinia amylovora
C:Species: Erwinia amylovora
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T08471
R:Laby, R.J.; Kim, J.F.; Beer, S.V.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z16433
A:Accession: T08471
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-403 <LAB>
A:Cross-references: EMBL:M92994; NID:g4309677; PID:g4309678
A:Experimental source: strain 321
C:Genetics:
A:Gene: hrpN

Query Match 91.2%; Score 2050; DB 2; Length 403;
Best Local Similarity 99.0%; Pred. No. 6.7e-123;
Matches 399; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	35	MSLNTSGLCASTMQISIGGAGGNNLLGTSRQAGLGGNSALGLGGNQNDTVNQLAGLL	94
Db	1	MSLNTSGLCASTMQISIGGAGGNNLLGTSRQAGLGGNSALGLGGNQNDTVNQLAGLL	60
Qy	95	TGMMMMMSMGGGLMGGGLGGGLGNGLGGGGLGGLSNALNDMLGGSNTLIGSKGNN	154
Db	61	TGMMMMMSMGGGLMGGGLGGGLGNGLGGGGLGGLSNALNDMLGGSNTLIGSKGNN	120
Qy	155	TTSTTNSPLDQALGINSTSQNDSTSGTSDTSDDPMQQLKMFSEIMQSLFGDGODGT	214
Db	121	TTSTTNSPLDQALGINSTSQNDSTSGTSDTSDDPMQQLKMFSEIMQSLFGDGODGT	180
Qy	215	QSSSGGKOPTGEQNAKKYKGYTDALSGLMGNGLSQLLNGGLGGGGGAGTGDLGSSSL	274
Db	181	QSSSGGKOPTGEQNAKKYKGYTDALSGLMGNGLSQLLNGGLGGGGGAGTGDLGSSSL	240
Qy	275	GGKGLQNLSPDYDQGLNAVGTGIGMKAGIQALNDIGTHRHSSSTRSFVNKGDRAMAKEI	334
Db	241	GGKGLQNLSPDYDQGLNAVGTGIGMKAGIQALNDIGTHSDSSTRSFVNKGDRAMAKEI	300
Qy	335	GOFMDQYPEVFGKPOYKQPGQEVKTDKSWAKALSKPDGDMTPASMEQFNKAKMIKR	394
Db	301	GOFMDQYPEVFGKPOYKQPGQEVKTDKSWAKALSKPDGDMTPASMEQFNKAKMIKS	360
Qy	395	PMAGDTGNLQARGAGSSSLGIDAMMAGDAINNALGKLGAA	437
Db	361	AMAGDTGNLQARGAGSSSLGIDAMMAGDAINNALGKLGAA	403

RESULT 2

D70907

hypothetical glycine-rich protein Rv3514 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70907
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: D70807
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1489 <COL>
A:Cross-references: GB:AL02022; GB:AL123456; NID:g3261554; PIDN:CAA17751.1; PID:g292445
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3514
C:Superfamily: collagen alpha 1(IV) chain

Query Match 11.8%; Score 264.5; DB 2; Length 1489;
Best Local Similarity 26.1%; Pred. NO. 5e-09;
Matches 116; Conservative 27; Mismatches 149; Indels 153; Gaps 16;

QY 37 LNTSLGASTMQI-----SISGAGGNGL-----LGTSTRON 67
DB 955 LNTDGLSSATSGTCTGCTGCTGCTGCGAGDDBSAGTGTGGAGNAGAGGLANTGGTAGN 1014
QY 68 AGLGGNSALGGLGNDTVNOL-----AGLLTMMMMMSMMGGGLGGLGGLNG 121
DB 1015 AGIGDGGGG--GNGGQDGGSLGGLGQPGFAGGAGGKGGAGGAGGAGGAGGAG--G 1072
QY 122 LCGSGGLCEGLSNALNDMLGGS--LNTLGSKGNNTTSTNSPLDQALGINTSQNDST 179
DB 1073 QGGAGGAGISFSNSNGTGTGTGGTGGTGGN-----ACTGACDPCKCGT 1119
QY 180 SCTDSTSDSPMQQLKMFSEIMOSLFEGDGDGTGGSS--SGGKOPTGEQNAKKGVTD 238
DB 1120 GCTGCTGCGS-----GAGSGGANFNGTGTGTGTGTGTGGTGGKGG 1159
QY 239 ALSGLMGNLSQLGLNGGLG-----GGGGNAGTGLDSSLLGKGLQNLGSPVDYQQ 290
DB 1160 GIADGGPGGCG--GNAGVGKGTNGNGSGGTGTGTGGAGGAGGAGGLANTGGTA--- 1213
QY 291 LCNAGVTCIGKAGIOALNDICTHRHSSTRFVNKGDRAKAKEICQFMDQYPEVECKPQY 350
DB 1214 -GNA---GIG-----GDDGGGGNGGQ--GDSGSLGCGQPGF 1243
QY 351 QKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMKRPMDGTGNGNLQARGA 410
DB 1244 AGGPGG-----KGGAGNAGGTGGNGSCA 1267
QY 411 GSSSLGIDAMMAGDAINNMALGKL 435
DB 1368 GAGGQGGAGGAGISFSNGSGGT 1292

RESULT 3

B70523
hypothetical glycine-rich protein Rv0297 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jan-2000
C:Accession: B70523
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70523

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-591 <COL>
A:Cross-references: GB:Z96800; GB:AL123456; NID:g3261800; PIDN:CAB09596.1; PID:632165
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv0297
C:Superfamily: unassigned collagens

Query Match 11.3%; Score 255; DB 2; Length 591;
Best Local Similarity 26.8%; Pred. NO. 6.9e-09;
Matches 123; Conservative 24; Mismatches 168; Indels 144; Gaps 20;

QY 41 GLGASTMOTISIGGAGNGLLTSRONAGLGCNSALGLGGGNDNVNQLAGLLTGMM 100
DB 141 GNGSGAPQAGGAGGAAGFFG---NNGNGDGAGANGG-----AGGTAGWFFG 187
QY 101 MSMMCGGCLMG-----CGLCGGGLN-----GLGSGGLGEGLSNALNDML----- 140
DB 188 FGGNGGAGGIVAGINGLGGAGGDDGNAGFFGNGGNGMGAGACVNAVNPCLATPVT 247
QY 141 -----GSLNTLGSKG--GNNTTSTNSPLDQALGINTSQNDSTSG-----TDSTSDS 188
DB 248 PRANGNGNLNVCPCTAGCGDANGSAIGQAGAGGAGGAGGAGGAGGAGGAGGAG 307
QY 189 SDPMOQLLKMFEIMOSLFEGDGDGTGGSSSGKOPTGEQNAKKGVTDALSLMGNL 248
DB 308 G-----GAGDGAFFGNGGNGSGVEHT-----GATGS--SASGNGA 342
QY 249 SOLLNNGCLG-----GCGGNGAGTGLDGLSGKGLQNLGSPV 286
DB 343 TG--GNGGVGAPGAGGNGGHVSGSVNTAGAGKGGNGGTGGAGGP--GCHGGSVLSPV 399
QY 287 DXQQGLNAVGTGIGKAGIQALNDICTHRHSSTRS----FVNKGDRAKAKEICQFMDQY 342
DB 400 GDSGNGGAGDGD--GAGVSATDICTGGRGNGGHHGLWNGGCGGAGVG----- 449
QY 343 EVFGKPYQKPGQEVKTDKSWAKALSKPDDGMPASMEQFNKAKGMKRPMD----AG 398
DB 450 -----GVG-----GAGNAGAIGGHHGDDG-----GSYNTFIGSSEAG 480
QY 399 DTGNGNLARGAGGSSGLGDAMMAGDAINNMALGKLCA 437
DB 481 DGGKGLGGDGGGRGIFG--QFGAGGAGGAGGAGGAGGAGGA 517

RESULT 4

C70974
hypothetical glycine-rich protein Rv3388 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: C70974
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70974
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-731 <COL>
A:Cross-references: GB:AL009198; GB:AL123456; NID:g3242362; PIDN:CAA15773.1; PID:g266
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3388
C:Superfamily: elastin

Query Match 11.1%; Score 249; DB 2; Length 731;
Best Local Similarity 28.7%; Pred. NO. 2.1e-08;
Matches 96; Conservative 23; Mismatches 101; Indels 114; Gaps 15;

A:Accession: B70807
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1079 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17749.1; PID:g292444
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3512
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 11.0%; Score 247.5; DB 2; Length 1079;
Best Local Similarity 26.0%; Pred. NO. 4.2e-08;
Matches 119; Conservative 26; Mismatches 170; Indels 143; Gaps 16;

QY 32 NSRSLNTSLGLASTMQLTISIGGAGNGNLLGTSRONAGLGGNSALGLGGGNNDTVNOLA 91
DB 629 NRSNGTGGAGCGCGGANGAGGAGGAGSGGGTGGNGAGGAGGAGGAGGNGGTGNGGN 688
QY 92 GLLTGCMMMMSMMGGGL--MGGGLGGLGNGLGGSGGLGEGLSNALNDMLGGSLLNT--- 146
DB 689 G-----GNGGIAGMGGGAGTGSNGGNGGSG---GNGGNAGMGGNSGTGSG 733
QY 147 LSKCCNNTTSTNSPLDQALGINST-----SQNDSTSTSTSDSDSDPMQQL 196
DB 734 DGGAGGGAAGTGTGTGGG-GLTGTGTGGSGGTGGDGGNGGNGADNTAN----- 783
QY 197 KMFSEIMQSLFGDGDGTGGSSG-----GKQTEGEQNAKKG----- 235
DB 784 -----MTAAGCGDGGGDDGGFGGAGAGCGGLTAGANGTGGCGGAGCGGNCATCGCHP 838
QY 236 VTDALSLGLMGNLSOLLNGGLG-----GGGGNAGTGLDGSLLSGKKG 278
DB 839 LTDDPGNGGTG-----GNGGTGTGGAGIGSLGGGTGGDGGNGGNGGTGGGEGVGGAG 893
QY 279 LQNLSPDYDQOLNNAVGTGCMKACIOALNDICGTHRSSTRSFVNKGDRAMAKEIQPM 338
DB 894 -----GTGAAGNGDGGTGTGGTGGDGGAGGTGTGTGTGGLGD----- 931
QY 339 DQYEVFGKPOYKPGQEVKTTDDKSWAKALSKPDDGNTFASMEQFNKAKMKRPMAG 398
DB 932 ---PRVGG-----SGDGGTGGSGGAAGNGGNG-----GNAG 960
QY 399 DTGNGNTQARGAGGSSGLGIDAMMAGDAINNMLGKLGA 436
DB 961 AGCNGNGTGCAGG-IGCTGCGNGDAEPGVPFGAGGA 996

RESULT 8
F70971
hypothetical glycine-rich protein Rv3367 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 18-Feb-2000
C:Accession: F70971
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70971
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-588 <COL>
A:Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15752.1; PID:el20228
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3367
C:Superfamily: unassigned collagens

Query Match 11.0%; Score 247; DB 2; Length 588;
Best Local Similarity 30.4%; Pred. No. 2.2e-08;
Matches 117; Conservative 16; Mismatches 134; Indels 98; Gaps 21;

QY 52 GGAGG--NNGLLGTSRONAGLGGNSALGLGGGNNDTVNOLAGLLTGMMMMSMMGGGL 109
DB 283 GGAGGDANGNPANTSIANAGAGGNGAAG-GDGGANGAGGAGGAAAGAGSSVGGDGGNGG 341
QY 110 MGQ-----GLCGGLGNGLGGSGGLGEGLSNALNDMLGGSLLNTLTKSGKNNTTSTNSPLDQ 165
DB 342 AGGTGTNGHAGGAG-GAGGAGGRRGWLVG--NGGNGGN-----GAAGNGAIGGTGG---- 390
QY 166 ALGINSTQNDDSTSTSTSDSDPMQQLLMFSEIMQSLFGDGDGTGGSSGCKQPT 225
DB 391 AGCV-PANOGGNSALCT-----DPVGGDGGDGGNGGTGG-----T 424
QY 226 EGEONAYKKGVTDALSG-LMNGLSOLLNGGLG--GGGGNAGTGLDGSLLGKGLQNL 282
DB 425 GRCGCGGSGGAGGASGLWLMNGNG--GNGCTGGSGVGGNGIGDGDAG-CGNATSTS 481
QY 283 SGPVDYDQOLNNAVGTGCMKAGTQALNDICGTHRSSTRSFVNKGDRAMAKEIQFMDQYP 342
DB 482 SIFPDAGH-GNG--GAGGDAG-----HGTTGGDGGDHAGTGGRGLL----- 522
QY 343 EYFCKPOYKPGQEVKTTDDKSWAKALSKPDDGCMTPASMEQFNKAKMKRPMAGDTGN 402
DB 523 -----AGOHANSNGGGGGGTGGAGGTGHTG-TPGSGN-----AGGTGT 557
QY 403 GNLQARGAGGSSGLGIDAMMAGDAIN 427
DB 558 GNADSTNGGPGSDG-----LCGDAPN 578

RESULT 9
H70846
hypothetical glycine-rich protein Rv3345c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70846
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70846
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1538 <COL>
A:Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PIDN:CAA17117.1; PID:g326
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3345c
C:Superfamily: collagen alpha 1(IV) chain

Query Match 10.9%; Score 246; DB 2; Length 1538;
Best Local Similarity 27.2%; Pred. NO. 7.8e-08;
Matches 124; Conservative 32; Mismatches 168; Indels 132; Gaps 21;

QY 41 GLGASTMQISIGAGGNGNLLGTSRONAGLGG-----NSALGLGGGNNDTVN----- 88
DB 249 GTC-----GIGGNGDAGLFG---NGGAGGAGAGLPCAAGLNGDGDGSDGNGGTGGN 298
QY 89 -QLAGLLTGMMMMSMMGGGLMG-----GCLGGGLGN-GLGSGG 127
DB 299 GGRGGLLVG-----NGGAGGAGGVDGGGKGGAGDPFSAVNNAGGNGHGGNPGVGAGG 354
QY 128 LGEGLSNALNDMLGGSLLNTLGSKGNNTTSTNSPLDQALGINSTSQNDSTSTSTSD 187
DB 355 AG-GLLAGAHGAACATPTSCGCGCGGICATANSPL-QAGGACGNG-GHGLVNCGTGC 411
QY 188 SSDPMQQLLMFSEIMQSLFGDGDGTGGSSGCKQPTGEONAYKKGVTDALSLGLMNG 247
DB 412 AG-----GAGHAGSTGATGATLQPTGGNGT---NGGAGGHHGGNGG 450

A:Accession: F70963
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-778 <COL>
A:Cross-references: GB:280225; GB:AL123456; NID:g3242265; PIDN:CAB02341.1; PID:e266390;
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv2634c
C:Superfamily: unassigned collagens

Query Match 10.8%; Score 242; DB 2; Length 778;
Best Local Similarity 30.1%; Pred. NO. 6.4e-08;
Matches 91; Conservative 19; Mismatches 110; Indels 82; Gaps 14;

QY 41 GLGASTMQISIGGAG--GNNGLLGTSRQNALGNSALGLGG-----GNQNDTVNQLAGLL 94
DB 353 CNGGNANPFGSGAGCGGCTGLACTNGVNPSTANPTNCATDNSGNCQT----- 404
QY 95 TGMNMMMSMMGGGLMGGLGGGLG--NGLGGSGGLGEGLSNALNDMLGGSNTLGSKGG 152
DB 405 -----GGNG--GPGFAGGVBAGGVGGGGLGESLDG--NDGTGG-----KGG 443
QY 153 NNTTSTNPLDQALCINSTQNDSTSGTSDTSDSDPMDQOALLKMFSEIMQSLFQDCQD 212
DB 444 AGGTAGTGGAGGAGGAGGAGGIGTDCSAGV-----ATGGEAGD 481
QY 213 GTQSSSSGKGQPTGE-QNAYKKGVYTDAL---SGLMNG--LSQLLNGGLGG----- 259
DB 482 CATGVGAGVGCACGKGGCGGCHTVCVDATGCGDGGICGDCNGALGAAGCGCTGCACGNGG 541
QY 260 -----GOGGNAGTGLDGSLSLGGKQLNLSPV---DYQOLGNNAVGTGCMKAGIQALND 310
DB 542 RGMMLIGNGAGGAGGTGTGT--GGGGAAGFAGGVGGAGGSELTDGACTAEGTGGGLGGLGG 600
QY 311 IC 312
DB 601 VG 602

RESULT 13
D70835
hypothetical glycine-rich protein Rv0278c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70835
R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: D70835
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-957 <COL>
A:Cross-references: GB:AL021930; GB:AL123456; NID:g3261524; PIDN:CAA17353.1; PID:g290946
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv0278c
C:Superfamily: elastin

Query Match 10.7%; Score 241.5; DB 2; Length 957;
Best Local Similarity 25.8%; Pred. NO. 8.7e-08;
Matches 117; Conservative 42; Mismatches 149; Indels 145; Gaps 21;

QY 36 SLNTSGLGASTMQISIGGAGGNNLLGTSRQNALGGLG-----NSALGGLGGNQNDTVNQLA 91
DB 362 TLTAGGICGA-----GAGGNAGLLGLGGGGGAGGFGFADGGGQGGCN-----A 407
QY 92 GLLTGMMMMMSMMGGGLMG-CGLGGGLGNGLGGSGGL-----CEGLSNAL---N 137
DB 408 G-----TVFGSGGAGGNGVGGFAGGIGGAGGTGGLLNGGNGGNGGASVATG 458

QY 138 DMLGSLNTLGSKGNNTTSTNSPLDQALGINSTQNDSTSGTSDTSDSDPMQQLK 197
DB 459 GGIGCTCVLICN--CGNGSGGIGAGKAGVGGVGLLLCLDGFNAPASTPSPLHTLQNNVLN 517
QY 198 MFSEIMQS-----LFGDGGDGTGGS-----SSGGKQPTGEGONAYKKGYTDALSG-LMNG 247
DB 518 VVNEPFTLTGRPLIGNANGTPTGTGADGGAGWLFNGANGTPTGTGAAGGAGWLFNG 577
QY 248 LSQLLGNGGLGG-----GOGNAGTGLDGSLSLGGKGLONLSGPVDY 288
DB 578 -----GNGGHGATNTATATGAGGAGGILFTGTGGNGTG--GIATGAGGIGGAGGAGV 630
QY 289 QOLGNNAVGTG-----ICMKAGIQALNDICTHRSSTRSFVNKQKDRAMAKEIGQFMDQVP 342
DB 631 SLLIGSGTGGNGSGISGV-AGIGGAGGRG-----GDAGL----- 664
QY 343 EYFGKPYQYQPGQEVKTDDBKNAKALSKPDDGCTFASHEQFNKAKMIKRPMDGTGN 402
DB 665 -LFG-----AACTGQ-----HCAAGGVDPAGVCCACGN 690
QY 403 GNLOARCAGSSSLGIDAMMAGDAINNMLKGLG 435
DB 691 GGLFANGAGGAGGFNA-AGGNGGNGGLFTGTG 722

RESULT 14
H70663
hypothetical glycine-rich protein Rv1840c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70663
R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70663
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-515 <COL>
A:Cross-references: GB:73859; GB:AL123456; NID:g3261678; PIDN:CAB06114.1; PID:gl7812
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv1840c
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 10.6%; Score 238.5; DB 2; Length 515;
Best Local Similarity 27.3%; Pred. NO. 6.6e-08;
Matches 109; Conservative 23; Mismatches 128; Indels 139; Gaps 15;

QY 52 GCAGGNNGLLGTSRQNALGNSALGGLGGNQNDTVNQLAGLLTGMNMMMSMMGGGLMG 111
DB 148 GGAGGDAGLIGNG-GNGGIGGPGATGLAGG-----AGGVGGLLFGDGGNGGAGGLG 197
QY 112 GGLGGGLNGLGGSGGLGEGLSNALNDMLGGSNTLGSKGNNTTSTNSPL--DQALGI 169
DB 198 TGPVGATG-CIGGPGGAAVGLPGHGGAGGAGGLCKAGFACGACGCTGTGGLLYGNGGNGG 256
QY 170 NSTSQNDSTSGTSDTSDSDPMDQOALLKMFSEIMQSLFQDGGDGTGQSSSGGKQPTGEQ 229
DB 257 NYFSGAADGGAGGDA-----RLINGNGDG-----GSVGAAPTG--- 289
QY 230 NAYKKGVTDALSGLMGNLSQLLNGGLGQ-----GQGNAGTGLDGSLSLGGKGLONLSGPV 286
DB 290 -----ICNG-----GNGNGWLYGDSGGSTLQGSDDGTG----- 322
QY 287 DYQOLGNNAVGTGIMKAGIQALNDICTHRSSTRSFVNKQKDRAMAKEIGQFMDQYPEVFG 346
DB 323 -----GNAGMFGGNGGFSFPDNGGDCGTGCTLTCNGCD----- 358

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:26:41 ; Search time 13 Seconds
(without alignments)

678.310 Million cell updates/sec

Title: 693_ll_3

Perfect score: 2248

Sequence: 1 MGFFLSQMPFFLVSTLL.....DAMMAGDAINNALGKLGAA 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 20178551 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2079	92.5	403	10	US-09-086-118-23
2	2079	92.5	403	10	US-09-835-684-3
3	2079	92.5	403	10	US-09-880-371-3
4	2079	92.5	403	10	US-09-879-248-3
5	2079	92.5	403	10	US-09-770-693-3
6	2079	92.5	403	10	US-09-766-348-3
7	718.5	32.0	338	10	US-09-086-118-21
8	718.5	32.0	338	10	US-09-835-684-1
9	718.5	32.0	338	10	US-09-880-371-1
10	718.5	32.0	338	10	US-09-879-248-1
11	718.5	32.0	338	10	US-09-770-693-1
12	718.5	32.0	338	10	US-09-766-348-1
13	211.5	9.4	651	10	US-09-861-597-1
14	198	8.8	606	10	US-09-861-597-6
15	197.5	8.8	344	10	US-09-086-118-27
16	197.5	8.8	344	10	US-09-835-684-11
17	197.5	8.8	344	10	US-09-880-371-11
18	197.5	8.8	344	10	US-09-879-248-15
19	197.5	8.8	344	10	US-09-770-693-7

20	197.5	8.8	344	10	US-09-766-348-7	Sequence 7, Appli
21	197.5	8.8	357	10	US-09-864-761-35807	Sequence 35807, A
22	197	8.8	606	10	US-09-861-597-8	Sequence 8, Appli
23	192.5	8.6	606	10	US-09-861-597-4	Sequence 4, Appli
24	188	8.4	283	10	US-09-864-761-36720	Sequence 36720, A
25	174.5	7.8	440	9	US-10-066-500-106	Sequence 106, App
26	174.5	7.8	440	9	US-10-063-547-52	Sequence 52, Appl
27	174.5	7.8	440	9	US-10-174-590-202	Sequence 202, App
28	174.5	7.8	440	9	US-10-176-758-202	Sequence 202, App
29	174.5	7.8	440	9	US-10-063-616-52	Sequence 52, Appl
30	174.5	7.8	440	9	US-10-175-737-202	Sequence 202, App
31	174.5	7.8	440	12	US-10-006-867-52	Sequence 52, Appl
32	174.5	7.8	440	12	US-10-052-586-202	Sequence 202, App
33	172	7.7	579	9	US-10-108-605-215	Sequence 215, App
34	169	7.5	34	10	US-09-770-693-11	Sequence 11, Appl
35	169	7.5	35	10	US-09-880-371-16	Sequence 16, Appl
36	165	7.3	34	10	US-09-770-693-13	Sequence 13, Appl
37	164.5	7.3	597	10	US-09-793-306-146	Sequence 146, Appl
38	163.5	7.3	479	10	US-09-918-951-3	Sequence 3, Appli
39	162.5	7.2	201	10	US-09-848-990-22	Sequence 22, Appl
40	162.5	7.2	201	10	US-09-760-364-14	Sequence 14, Appl
41	159.5	7.1	1230	10	US-09-881-752A-150	Sequence 150, App
42	158	7.0	943	9	US-09-996-634-131	Sequence 131, App
43	156	6.9	191	10	US-09-864-761-36985	Sequence 36985, A
44	154.5	6.9	447	10	US-09-835-684-5	Sequence 5, Appli
45	154.5	6.9	447	10	US-09-880-371-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-086-118-23
; Sequence 23, Application US/09086118
; Patent No. US20010011380A1

; GENERAL INFORMATION:

; APPLICANT: Laby, Ronald J.

; APPLICANT: Beer, Steven V.

; APPLICANT: Wei, Zhong-Min

; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR

; TITLE OF INVENTION: FRAGMENTS ELICITING A HYPERSENSITIVE RESPONSE AND USES

; TITLE OF INVENTION: THEREOF

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSES: Nixon, Hargrave, Devans & Doyle LLP

; STREET: Clinton Square, P.O. Box 1051

; CITY: Rochester

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 14603

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/086,118

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/048,109

; FILING DATE: 30-MAY-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Goldman, Michael L.

; REGISTRATION NUMBER: 30,727

; REFERENCE/DOCKET NUMBER: 19603/1301

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (716) 263-1304

; TELEFAX: (716) 263-1600

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 403 amino acids

; TYPE: amino acid


```
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Erwinia chrysanthemi
US-09-770-693-1

Query Match      32.0%; Score 718.5; DB 10; Length 338;
Best Local Similarity 42.8%; Pred. No. 3.4e-48;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 47 MOISI-GGAGGNNGLGTSRONAGLGG-NSA-----LGLCCGNQNDTVNQLAGLLTGMWMM 101
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 1 MOITIKAHITGGDLGVSLGAQ--GLKGLNSAASSLSSVDKLSSTIDKLTSAITSMW--- 55
QY 102 SMWGGGLMGGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 56 -----FGCALAQLGAS-SKGLCMSNQLGQSGFN-----GAQGNLLSVPK- 96
QY 162 PLDQALGINSTSONDDSTSGTSDSSDPMQOQLLKMFSEIMQSLFG-----DQDGC 213
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 97 -----SGGDALS-----KMFDRALDDLGHDTVTVKLTNQSNO 128
QY 214 TQSSSSGGKQPTGEQNAVYKGYTDALSLGMLNGLSQLLGGGQSGNAGTGLDGSS 273
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 129 LANSMLNASOMTQGNNAFGSGVNNALSSILGNGLGQSM-----SGFSQPS 174
QY 274 LGGKGLQNLSPVDYQOGLGNNAVGTGCMKAGIOALNDIGTHRRSHSTRSFYNKGDRAK 333
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 175 LGAGGLQCLSGAGAFNOLGNAIGMVGCONAALSALSNSVTHVDGNNRHFVDKEDRGMA 234
QY 334 IGQFMDOYPEVFGKPYQKPGQYKTDKSKWAKALSKPDDGDMTPASMEQFNKAKGM 393
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 235 IGQFMDOYPEIFGKPEYQKDGWSSPKTDDKSWAKALSKPDDGDMTGASMDKFRQAMG 294
QY 394 RPNAGDTGNLQARGAGSSSLGIDAMMAGDAINNMLGKLGAA 437
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 295 SAVAGDTGNTNLNRGAGGASLGIDAAVVGDKIANMSLGKLANA 338

RESULT 12
US-09-766-348-1
; Sequence 1, Application US/09766348
; Patent No. US20020116733A1
; GENERAL INFORMATION:
; APPLICANT: Qiu, Dewen
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED RESISTANCE IN PLANTS BY
; FILE REFERENCE: 19603/2986
; CURRENT APPLICATION NUMBER: US/09/766,348
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 08/984,207
; PRIOR FILING DATE: 1997-12-03
; PRIOR APPLICATION NUMBER: 60/033,230
; PRIOR FILING DATE: 1996-12-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Erwinia chrysanthemi
US-09-766-348-1

Query Match      32.0%; Score 718.5; DB 10; Length 338;
Best Local Similarity 42.8%; Pred. No. 3.4e-48;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 47 MOISI-GGAGGNNGLGTSRONAGLGG-NSA-----LGLCCGNQNDTVNQLAGLLTGMWMM 101
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 1 MOITIKAHITGGDLGVSLGAQ--GLKGLNSAASSLSSVDKLSSTIDKLTSAITSMW--- 55
QY 102 SMWGGGLMGGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 102 SMWGGGLMGGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
QY 102 SMWGGGLMGGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
QY 102 SMWGGGLMGGGLGGLGNGLSGGLGSLNSALNDMLGGSTNTLCSKGCNNNTTSTNS 161
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
```

```
Db 56 -----FGGALAQGLGAS-SKGLCMSNQLGQSGFN-----GAQGNLLSVPK- 96
QY 162 PLDQALGINSTSONDDSTSGTSDSSDPMQOQLLKMFSEIMQSLFG-----DQDGC 213
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 97 -----SGGDALS-----KMFDRALDDLGHDTVTVKLTNQSNO 128
QY 214 TQSSSSGGKQPTGEQNAVYKGYTDALSLGMLNGLSQLLGGGQSGNAGTGLDGSS 273
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 129 LANSMLNASOMTQGNNAFGSGVNNALSSILGNGLGQSM-----SGFSQPS 174
QY 274 LGGKGLQNLSPVDYQOGLGNNAVGTGCMKAGIOALNDIGTHRRSHSTRSFYNKGDRAK 333
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 175 LGAGGLQCLSGAGAFNOLGNAIGMVGCONAALSALSNSVTHVDGNNRHFVDKEDRGMA 234
QY 334 IGQFMDOYPEVFGKPYQKPGQYKTDKSKWAKALSKPDDGDMTPASMEQFNKAKGM 393
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 235 IGQFMDOYPEIFGKPEYQKDGWSSPKTDDKSWAKALSKPDDGDMTGASMDKFRQAMG 294
QY 394 RPNAGDTGNLQARGAGSSSLGIDAMMAGDAINNMLGKLGAA 437
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 295 SAVAGDTGNTNLNRGAGGASLGIDAAVVGDKIANMSLGKLANA 338

RESULT 13
US-09-861-597-1
; Sequence 1, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861,597
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: PR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-861-597-1

Query Match      9.4%; Score 211.5; DB 10; Length 651;
Best Local Similarity 25.5%; Pred. No. 6.6e-09;
Matches 108; Conservative 21; Mismatches 168; Indels 127; Gaps 17;

QY 41 GLCASTMOISIGGAGGNNGLLGTSRQNAIGLGNLSAL-----GLGGGNNQNDTVNQLAGL-- 93
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 104 GGAGAGAAAAAAGAGAGGGGGYGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 163
QY 94 LTCMMHMMHMMGGGGLMCGGLG--GGLCNGLCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 151
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 164 LGGGAGAGAAAAAGAGGGGGYGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 214
QY 152 GNNTTSTNSPLDQ-ALGINSTSONDDSTSGTSDSSDPMQOQLLKMFSEIMQSLFGDC 210
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 215 AGAAAAAAGAGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 256
QY 211 QDGTGSSSSGGKQPTGEQNAVYKGYTDALSLGMLNGLSQLLGGGQSGNAGTGLDGSS 267
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 257 GLSSQAGAGGG-----EG-----AGAAAAAGAGAGCG-----GYGGLCGGAGCGGCGGL 301
QY 268 GLDSSSLCGCKGLON-----LSCPVDVVOOLCHNAVCTGICMKAGIOALNDIGTHRRSHSTRSF 323
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
D 302 GSQAGAGGGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 342
|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
```


QY 324 NKGDRAKEIGOFMDQYPEYFGKPOYKQPGQEVKTTDDKSWAKALSKPDDDDGWT PASME 383
Db 343 -----AAAGGAGG--GGYGGGSGCAGRGGLG----- 367
QY 384 OFNKAQMIKRPMDAGDTGNGNL-----QARGAGSSGLIDAMMAGDA-----INNMA 431
Db 368 --GOGAGAVAAAAGGAGGCGGYGGLGSGCAGRGCGAGCAAAAAGGAGGCGRGYGGGLGNOGA 425
QY 432 GKLK 435
Db 426 GRGG 429
RESULT 14
US-09-861-597-6
; Sequence 6, Application US/09861597
; Patent No. US20020064539A1
; GENERAL INFORMATION:
; APPLICANT: PHILLIPPE, Michel
; APPLICANT: GARSON, Jean-Claude
; APPLICANT: ARRAUDEAU, Jean-Pierre
; TITLE OF INVENTION: COSMETIC OR DERMATOLOGICAL COMPOSITION CONTACTING AT
; TITLE OF INVENTION: LEAST ONE NATURAL OR RECOMBINANT SPIDER SILK OR AN
; TITLE OF INVENTION: ANALOG
; FILE REFERENCE: 6388-0365-0
; CURRENT APPLICATION NUMBER: US/09/861.597
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/247,806
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: FR 98/01614
; PRIOR FILING DATE: 1998-02-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 606
; TYPE: PRT
; ORGANISM: Nephila clavipes
US-09-861-597-6

Query Match 8.8%; Score 198; DB 10; Length 606;
Best Local Similarity 25.2%; Pred. No. 6.6e-08;
Matches 114; Conservative 21; Mismatches 182; Indels 136; Gaps 17;
QY 41 GLGASTMQISTGGAG-GNNGLLGTSRQWAGLGGNSALGLGGGNDTNNQLAGLLTGM 99
Db 180 GOGAGAAAAGGAGCGGYGLGS--QGAGCGGYGGLGSGCAGRGGLGCGGAGAAAAA 237
QY 100 MMSMMGGGLMGGLG-----GGLGNGLGGSGGLGGLGSLNALNDMLG-- 141
Db 238 GGAGGGLGSGCAGGAGAAAAAGGAGGCGGYGGLGSGCAGRGCGGAGAAAAAGGAGCG 297
QY 142 -----GSLNTLGSKG-GNNTTSTNSPLDQALGINSTSQNDSTSGTSDS 188
Db 298 GYGGLGSGCAGGCGGYGGLGSGCAGRGGLGSGGAGAAAAAGGAGCGGAGGAGAGA 357
QY 189 SDPMQQLKMFSEIMQSLFSGDQGTGSSSGSKOPTGEQNAKKGYTDLGSLMGNGL 248
Db 358 A-----AAAAGGAGCGGYGGLGSGQ--AGRGGGA---GAAAAAGGAGG- 398
QY 249 SOLLNGGL---GGGQGNAGTGLDSSSLGKGLQN-----LSGPVDYQQLGN-AVGT 297
Db 399 ---GYGGLGSGCAGGCGGYGGLGSGCAGRGGLGSGCAGAAAAAGGAGCGGAGGAG 454
QY 298 GIGMKA-----GTQALNDITGTHRSSTRSFVNKGDRAKEMKIGFMDQYPEVFKP 348
Db 455 GAGAAAAAGGAGCGGYGGLGSGCAGRG-----GCGAGAAAAAG----- 494
QY 349 QYKQGGGVKTTDDKSWAKALSKPDDDGWT PASMEQFNKAKGMIKRPMDGTGNCNLQAR 408
Db 495 ---GAGQ-----GGSGLGIDAMMAGDAINNMA 436
QY 409 GA-----GGSGLGIDAMMAGDAINNMA 436

Db 521 GAGRGGLGCGGAGAAAAAAGGAGCGGLGSGGA 553
RESULT 15
US-09-086-118-27
; Sequence 27, Application US/09086118
; Patent No. US20010011380A1
; GENERAL INFORMATION:
; APPLICANT: Laby, Ronald J.
; APPLICANT: Beer, Steven V.
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR
; TITLE OF INVENTION: FRAGMENTS ELICITING A HYPERSENSITIVE RESPONSE AND USES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/086,118
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,109
; FILING DATE: 30-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1301
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-086-118-27

Query Match 8.8%; Score 197.5; DB 10; Length 344;
Best Local Similarity 26.7%; Pred. No. 3.7e-08;
Matches 75; Conservative 31; Mismatches 86; Indels 89; Gaps 12;
QY 55 GGNNGLLGTSRQWAGLGG-NSALGLGGGNDTNNQLAGLLTGMMSMMGGGLMG 113
Db 136 GGN-----KNGVGGGANGKAGG-----OGGLAEALQIEQILAQLGCGGAGG 182
QY 114 LGGGLNGLGGSGGLGGLSALNDMLGSLNTLGSKGGNNTTSTNSPLDQALGINST 173
Db 183 AGGVGGAGGADGGSGAGGAGGAN-----GADGGN-----GVNGNQ 218
QY 174 QNDSTSGTSDTSDDSPMQQLKMFSEIMQSLFGDQGTGSSSGGKOPTGEQNAK 233
Db 219 ANGPNQAGDVNGANGAD-----DCSEDDG----- 242
QY 234 KGVTDALSLMG---NGLSOLLNGGLGG---QGGNAGTGLDSSSLGKGLQNLGSPVDY 288
Db 243 -GLTVGLQKMLKILNALVQMMQGGGLGGNQAGGSKGAGNAPASGANPNQPGSADD 301
QY 289 QQLG-NAVGTGI--GMKAGTQALNDI-----CTHRHSSTR 320
Db 302 QSSGQNNLQSLINDVYKVVQILQMLAAQNGGSSQSTSTQ 342

Search completed: January 22, 2003, 16:29:12
Job time : 14 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:26:07 ; Search time 18 seconds
(without alignments)
714.323 Million cell updates/sec

Title: 693_11_3

Perfect score: 2248

Sequence: 1 MGFFLFQSPSFVFLVSTLL.....DAMMAGDAINNALGKLGA 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2079	92.5	403	2	US-08-200-724A-2
2	2079	92.5	403	2	US-09-030-270A-3
3	2079	92.5	403	4	US-08-851-376A-2
4	2079	92.5	403	4	US-08-984-207-3
5	2079	92.5	403	4	US-09-013-587-3
6	1928	85.8	385	1	US-08-891-254-3
7	1928	85.8	385	5	PCT-US96-08819-3
8	1928	85.8	385	5	PCT-US96-08819-3
9	1913	85.1	385	1	US-08-891-254-3
10	718.5	32.0	338	1	US-08-891-254-1
11	718.5	32.0	338	2	US-08-484-358-2
12	718.5	32.0	338	2	US-08-819-539-1
13	718.5	32.0	338	2	US-09-030-270A-1
14	718.5	32.0	338	3	US-09-118-959-2
15	718.5	32.0	338	4	US-08-984-207-1
16	718.5	32.0	338	4	US-09-013-587-1
17	718.5	32.0	338	5	PCT-US96-08819-1
18	211.5	9.4	651	4	US-09-060-756-727
19	211.5	9.4	651	4	US-08-556-978B-19
20	211.5	9.4	651	4	US-09-247-806-1
21	211.5	9.4	718	1	US-08-425-069-2
22	211.5	9.4	718	2	US-08-317-844B-2
23	211.5	9.4	747	3	US-09-034-177-3
24	206	9.2	604	4	US-08-556-978B-63
25	200	8.9	334	4	US-09-060-756-728
26	198.5	8.8	738	3	US-08-864-038A-3
27	198	8.8	606	4	US-09-247-806-6

28	197.5	8.8	344	1	US-08-891-254-7	Sequence 7, Appli
29	197.5	8.8	344	2	US-08-819-539-7	Sequence 7, Appli
30	197.5	8.8	344	2	US-09-030-270A-7	Sequence 7, Appli
31	197.5	8.8	344	4	US-08-984-207-7	Sequence 7, Appli
32	197.5	8.8	344	4	US-09-013-587-7	Sequence 7, Appli
33	197.5	8.8	344	5	PCT-US96-08819-7	Sequence 23, Appli
34	197	8.8	606	4	US-08-556-978B-23	Sequence 8, Appli
35	197	8.8	606	4	US-09-247-806-8	Sequence 24, Appli
36	196	8.7	1160	3	US-08-808-599A-24	Sequence 21, Appli
37	192.5	8.6	606	4	US-08-556-978B-21	Sequence 4, Appli
38	192.5	8.6	606	4	US-09-247-806-4	Sequence 3, Appli
39	190	8.5	641	4	US-09-249-585A-3	Sequence 2, Appli
40	187.5	8.3	749	1	US-08-317-522A-2	Sequence 2, Appli
41	187.5	8.3	749	1	US-08-439-818A-2	Sequence 2, Appli
42	187.5	8.3	749	2	US-08-751-965-2	Sequence 2, Appli
43	187.5	8.3	749	2	US-08-738-975-2	Sequence 2, Appli
44	187.5	8.3	749	2	US-08-728-626-2	Sequence 2, Appli
45	187.5	8.3	749	3	US-08-808-599A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-200-724A-2
; Sequence 2, Application US/08200724A
; Patent No. 5849868
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Bauer, David W.
; APPLICANT: Beer, Steven V.
; APPLICANT: Collmer, Alan
; APPLICANT: He, Sheng-Yang
; APPLICANT: Laby, Ron J.
; TITLE OF INVENTION: ELICITOR OF THE HYPERSENSITIVE RESPONSE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,724A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/10030
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-200-724A-2

Query Match 92.5%; Score 2079; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 5,9e-176;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	92.5%	Score 2079;	DB 2;	Length 403;
Best Local Similarity	100.0%;	Prod. No. 5.9e-176;		
Matches 403;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps
35	MSINTSGTLCASPMWSTIGACGNCNCLIGTSSRONAGLGGNSALGIGGGGQNDVTYNQIAGLL	94		
1	MSINTSGUGASTWQLISIGAGGNCNCLIGTSSQNAGLGGNSALGIGGGGQNDVTYNQIAGLL	60		
95	TGMMMMMSMMGGGGTGGGGLGAGLNGTGGCGGTGEGLSNALNDMLGCSLNTTIGSKGGNN	154		
61	TGMMMMMSMMGGCGTGGGGLGAGLNGTGGCGGTGEGLSNALNDMLGCSLNTTIGSKGGNN	120		
155	TTTSTTNSPLDQALGINTSYNSQNDSTSGTSDSTSDSDSDPQQLIKMFSEIMQSLFCDDQDGT	214		
121	TTSTTNSPLDQALGINTSYNSQNDSTSGTSDSTSDSDSDPQQLIKMFSEIMQSLFCDDQDGT	180		
215	QGSSSGCKOPTGEGONAYKKGYTDALSGLMGNGLSQLLNGGGLGGGGGAGNAGTGLGGSSL	274		
181	QGSSSGCKOPTGEGONAYKKGYTDALSGLMGNGLSQLLNGGGLGGGGGAGNAGTGLGGSSL	240		
275	GKGGLQNTSGPVDYQQLGNAGVGTGTGMKAGTQALNDGTGTHRHSSSTRSFVNKGGRAMAKEI	334		
241	GKGGLQNTSGPVDYQQLGNAGVGTGTGMKAGTQALNDGTGTHRHSSSTRSFVNKGGRAMAKEI	300		
335	GQFMDOYPEVFCKPOYQKPGQEVKTDDBKSWAKALSKPDDGDMTPASMEQFNKAKGMIKR	394		
301	GQFMDOYPEVFCKPOYQKPGQEVKTDDBKSWAKALSKPDDGDMTPASMEQFNKAKGMIKR	360		
395	PMAGDTGNCNLQARGAGGSSLIGDAMMAGDAINNMLCKLGAA	437		
361	PMAGDTGNCNLQARGAGGSSLIGDAMMAGDAINNMLCKLGAA	403		

RESULT 3
US-08-851-376A-2
; Sequence 2, Application US/08851376A
; Patent No. 6174717
; GENERAL INFORMATION:
; APPLICANT: Beer, Steven V.
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Bauer, David W.
; APPLICANT: Collmer, Alan
; APPLICANT: He, Sheng-yang
; APPLICANT: Lady, Ron
; TITLE OF INVENTION: ELICITOR OF THE HYPERSENSITIVE RESPONSE
; TITLE OF INVENTION: IN PLANTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon Peabody LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,376A
; FILING DATE: 05-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/200,724
; FILING DATE: 23-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30737
; REFERENCE/DOCKET NUMBER: 19603/10035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304

TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-851-376A-2

Query Match 92.5%; Score 2079; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 5.9e-176;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 MSLNTSGLCASMTQISIGGAGNNGLLGTSRQNALGGSALGLGGNQNDTVNQLAGLL 94
Db 1 MSLNTSGLCASMTQISIGGAGNNGLLGTSRQNALGGSALGLGGNQNDTVNQLAGLL 60
QY 95 TGMMMMSMMGGGLMGGGLGNGLGGGGLGEGLSNALNDMLGGSNLTLGSKGNN 154
Db 61 TGMMMMSMMGGGLMGGGLGNGLGGGGLGEGLSNALNDMLGGSNLTLGSKGNN 120
QY 155 TTSTTNSPLDQALGINSQNDSTSGTSDSDPQQQLKMFSEIMQSLFSGDQDGT 214
Db 121 TTSTTNSPLDQALGINSQNDSTSGTSDSDPQQQLKMFSEIMQSLFSGDQDGT 180
QY 215 QGSSSGGKQPTGEGQYKKGVTDALSGLMGNGLSQLLGGGGLGGGAGTGLDGSLL 274
Db 181 QGSSSGGKQPTGEGQYKKGVTDALSGLMGNGLSQLLGGGGLGGGAGTGLDGSLL 240
QY 275 GKGGLNLGSPVDYQQLGNVAGTGTGIMKAGIQALNDIGTHRHSSYRFSVFNKGDRAKAI 334
Db 241 GKGGLNLGSPVDYQQLGNVAGTGTGIMKAGIQALNDIGTHRHSSYRFSVFNKGDRAKAI 300
QY 335 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAGMIKR 394
Db 301 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAGMIKR 360
QY 395 PMAGDTGNGNLQARGAGSSLGIDAMMAGDAINNMLGKLGA 437
Db 361 PMAGDTGNGNLQARGAGSSLGIDAMMAGDAINNMLGKLGA 403

RESULT 4
US-08-984-207-3
; Sequence 3, Application US/08984207
; Patent No. 6235974
; GENERAL INFORMATION:
; APPLICANT: Oiu, Dwen
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HIPERSENSITIVE RESPONSE INDUCED
; TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,207
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,230
; FILING DATE: 05-DEC-1996

ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1201
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 403 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-984-207-3

Query Match 92.5%; Score 2079; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 5.9e-176;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 MSLNTSGLCASMTQISIGGAGNNGLLGTSRQNALGGSALGLGGNQNDTVNQLAGLL 94
Db 1 MSLNTSGLCASMTQISIGGAGNNGLLGTSRQNALGGSALGLGGNQNDTVNQLAGLL 60
QY 95 TGMMMMSMMGGGLMGGGLGNGLGGGGLGEGLSNALNDMLGGSNLTLGSKGNN 154
Db 61 TGMMMMSMMGGGLMGGGLGNGLGGGGLGEGLSNALNDMLGGSNLTLGSKGNN 120
QY 155 TTSTTNSPLDQALGINSQNDSTSGTSDSDPQQQLKMFSEIMQSLFSGDQDGT 214
Db 121 TTSTTNSPLDQALGINSQNDSTSGTSDSDPQQQLKMFSEIMQSLFSGDQDGT 180
QY 215 QGSSSGGKQPTGEGQYKKGVTDALSGLMGNGLSQLLGGGGLGGGAGTGLDGSLL 274
Db 181 QGSSSGGKQPTGEGQYKKGVTDALSGLMGNGLSQLLGGGGLGGGAGTGLDGSLL 240
QY 275 GKGGLNLGSPVDYQQLGNVAGTGTGIMKAGIQALNDIGTHRHSSYRFSVFNKGDRAKAI 334
Db 241 GKGGLNLGSPVDYQQLGNVAGTGTGIMKAGIQALNDIGTHRHSSYRFSVFNKGDRAKAI 300
QY 335 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAGMIKR 394
Db 301 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAGMIKR 360
QY 395 PMAGDTGNGNLQARGAGSSLGIDAMMAGDAINNMLGKLGA 437
Db 361 PMAGDTGNGNLQARGAGSSLGIDAMMAGDAINNMLGKLGA 403

RESULT 5
US-09-013-587-3
; Sequence 3, Application US/09013587
; Patent No. 6277814
; GENERAL INFORMATION:
; APPLICANT: Oiu, Dwen
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: ENHANCEMENT OF GROWTH IN PLANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/013,587

```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/036,048
; FILING DATE: 27-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-013-587-3

Query Match          92.5%; Score 2079; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 5.9e-176;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 MSLNTSGLGASTMQISIGGAGGNNLLGTSRONAGLGGNSALGLGGGNNQNDTVYNQLAGLL 94
DB 1 MSLNTSGLGASTMQISIGGAGGNNLLGTSRONAGLGGNSALGLGGGNNQNDTVYNQLAGLL 60
QY 95 TGMWMMWMMWGGGLMGCGGLGGGLGNGLGCGGGLGEGLSNALNDMLGGSNTLGSKGGNN 154
DB 61 TGMWMMWMMWGGGLMGCGGLGGGLGNGLGCGGGLGEGLSNALNDMLGGSNTLGSKGGNN 120
QY 155 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLLKMFSFIMQSLFGDQDGT 214
DB 121 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLLKMFSFIMQSLFGDQDGT 180
QY 215 QGSSGGKOPTGEQONAYKKGVTDALSGLMGNGLSQLLGGNGGLGGGGAGTGLDGSSL 274
DB 181 QGSSGGKOPTGEQONAYKKGVTDALSGLMGNGLSQLLGGNGGLGGGGAGTGLDGSSL 240
QY 275 GKGGLONLSPGVYQOLGNVAGTGTGMKAGTQALNDIGTHRHSSSTRSFVNKGDRAKAI 334
DB 241 GKGGLONLSPGVYQOLGNVAGTGTGMKAGTQALNDIGTHRHSSSTRSFVNKGDRAKAI 300
QY 335 GQFMDQYPEVFGKPOYKQGPQGVKTDKSWAKALSKPDDDDGMTFASME2FNKAKGMIKR 394
DB 301 GQFMDQYPEVFGKPOYKQGPQGVKTDKSWAKALSKPDDDDGMTFASME2FNKAKGMIKR 360
QY 395 PMAGDTGNGNLQARAGAGSSSLGIDAMMAGDAINNHALGKLGA 437
DB 361 PMAGDTGNGNLQARAGAGSSSLGIDAMMAGDAINNHALGKLGA 403

RESULT 6
US-08-891-254-3
; Sequence 3, Application US/08891254
; Patent No. 5776889
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: Hypersensitive Response
; TITLE OF INVENTION: Induced Resistance In Plants
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,254
; FILING DATE: 10-JUL-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/475,775
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 14603/10050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 385 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-891-254-3

Query Match          85.8%; Score 1928; DB 1; Length 385;
Best Local Similarity 100.0%; Pred. No. 1.2e-162;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 MSLNTSGLGASTMQISIGGAGGNNLLGTSRONAGLGGNSALGLGGGNNQNDTVYNQLAGLL 94
DB 1 MSLNTSGLGASTMQISIGGAGGNNLLGTSRONAGLGGNSALGLGGGNNQNDTVYNQLAGLL 60
QY 95 TGMWMMWMMWGGGLMGCGGLGGGLGNGLGCGGGLGEGLSNALNDMLGGSNTLGSKGGNN 154
DB 61 TGMWMMWMMWGGGLMGCGGLGGGLGNGLGCGGGLGEGLSNALNDMLGGSNTLGSKGGNN 120
QY 155 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLLKMFSFIMQSLFGDQDGT 214
DB 121 TTSTTNSPLDQALGINSTQNDSTSGTSDSDSPMQQLLKMFSFIMQSLFGDQDGT 180
QY 215 QGSSGGKOPTGEQONAYKKGVTDALSGLMGNGLSQLLGGNGGLGGGGAGTGLDGSSL 274
DB 181 QGSSGGKOPTGEQONAYKKGVTDALSGLMGNGLSQLLGGNGGLGGGGAGTGLDGSSL 240
QY 275 GKGGLONLSPGVYQOLGNVAGTGTGMKAGTQALNDIGTHRHSSSTRSFVNKGDRAKAI 334
DB 241 GKGGLONLSPGVYQOLGNVAGTGTGMKAGTQALNDIGTHRHSSSTRSFVNKGDRAKAI 300
QY 335 GQFMDQYPEVFGKPOYKQGPQGVKTDKSWAKALSKPDDDDGMTFASMEQFNKAKGMIKR 394
DB 301 GQFMDQYPEVFGKPOYKQGPQGVKTDKSWAKALSKPDDDDGMTFASMEQFNKAKGMIKR 360
QY 395 PMAGDTGNGNLQ 406
DB 361 PMAGDTGNGNLQ 372

RESULT 7
US-08-819-539-3
; Sequence 3, Application US/08819539
; Patent No. 5859324
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: Hypersensitive Response
; TITLE OF INVENTION: Induced Resistance In Plants
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester

```



```

; TITLE OF INVENTION: Elicitor of the Hypersensitive Response in Plants
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates
; STREET: 25 skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06243
; FILING DATE: 19930630
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 907,935
; FILING DATE: 01-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: George M. Yahwak
; REGISTRATION NUMBER: 26,824
; REFERENCE/DOCKET NUMBER: CRF D-1172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203)268-1951
; TELEFAX: (203)268-1951
; INFORMATION FOR SEQ ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 385 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-06243-2

Query Match 85.1%; Score 1913; DB 5; Length 385;
Best Local Similarity 99.5%; Pred. No. 2.6e-161;
Matches 369; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 35 MSLNTSGLCASTMQISIGGAGGNNLLGTSRONAGLGGNSALGLGGGNQNDVTYNQLAGLL 94
Db 1 MSLNTSGLCASTMQISIGGAGGNNLLGTSRONAGLGGNSALGLGGGNQNDVTYNQLAGLL 60
QY 95 TGMNMMMSMMGGGLMGGGLGGLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLG 154
Db 61 TGMNMMMSMMGGGLMGGGLGGLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLGSLG 120
QY 155 TTSTTNSPLDQALGINSTSONDDSTSGTSDSDSPMQQLKMFSEIMQSLFSGDQDGT 214
Db 121 TTSTTNSPLDQALGINSTSONDDSTSGTSDSDSPMQQLKMFSEIMQSLFSGDQDGT 180
QY 215 QGSSSGGKQPTGEQNAKKYKGYTDALSLMGNLSOLLGNGLLGGGGGAGNAGTGLDGSLL 274
Db 181 QGSSSGGKQPTGEQNAKKYKGYTDALSLMGNLSOLLGNGLLGGGGGAGNAGTGLDGSLL 240
QY 275 GSKGLQNLGSPVDYQOOLGNAGVTGICMKAGIOALNDIGTHRHSSSTRSFYNKGDRAKKEI 334
Db 241 GSKGLQNLGSPVDYQOOLGNAGVTGICMKAGIOALNDIGTHRHSSSTRSFYNKGDRAKKEI 300
QY 335 GQFMDOYPEVFGKPYQKPGQEVKTDKSKWAKALSKPDDDGMTPASMEQFNKAKGMIR 394
Db 301 GQFMDOYPEVFGKPYQKPGQEVKTDKSKWAKALSKPDDDGMTPASMEQFNKAKGMIR 360
QY 395 PMAGDTGNGNL 405
Db 361 PMAGDTGNGNL 371

RESULT 10
US-08-891-254-1
g; Sequence 1, Application US/08891254
```

```

; Patent No. 5776889
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: Hypersensitive Response
; TITLE OF INVENTION: Induced Resistance In Plants
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,254
; FILING DATE: 10-JUL-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/475,775
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 14603/10050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO. 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-891-254-1

Query Match 32.0%; Score 718.5; DB 1; Length 338;
Best Local Similarity 42.8%; Pred. NO. 8.8e-56;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

QY 47 MQISI-GGAGGNNLLGTSRONAGLGG-NSA---LGLGGGNQNDVTYNQLAGLLTGMNMM 101
Db 1 MQITIKAHIGGDLGVSGLGAQ--GLKGLNSAASLGSVDKLSSTIDKLTSAITSMH--- 55
QY 102 SMMGGGGLMCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 161
Db 56 -----FGGALAQGLGAS-SKGLGMSNQLGQSGFN-----CAQASNLLSYPK - 96
QY 162 PLDQALGINSTSONDDSTSGTSDSDSPMQQLKMFSEIMQSLFSGDQDGT 213
Db 97 -----SGGDALS-----KMFADLDDLLGHDTVTIKLTNOSNQ 128
QY 214 TQSSSGGKQPTGEQNAKKYKGYTDALSLMGNLSOLLGNGLLGGGGGAGNAGTGLDGS 273
Db 129 LANSMLNASQMTQCNNAFCGVNNALSSILCNLCQSM-----SCFSPQS 174
QY 274 LGCKGLQNLGSPVDYQOOLGNAGVTGICMKAGIOALNDIGTHRHSSSTRSFYNKGDRAK 333
Db 175 LGAGGQGLGSGAGAFNQLGNAIGNVGQNAALSALSHVSTHVDGNNRHIFVDKEDRGWAK 234
QY 334 IQQFMDOYPEVFGKPYQKPGQEVKTDKSKWAKALSKPDDDGMTPASMEQFNKAKGMIR 393
Db 235 IQQFMDOYPEIFCKPEYQKDGWSSPKTDDKSWAKALSKPDDDGMTGASMDKFRQAMGMIR 294
QY 394 RPMAGDTGNGNLQARGAGGSLGIDAMMAGDAINNMLGKGLAA 437
Db 295 SAVAGDTGNTNLRLGAGGASLGIDAAVVGDKTANMSLCKLANA 338
```


RESULT 11
US-08-484-358-2
; Sequence 2, Application US/08484358
; Patent No. 5850015
; GENERAL INFORMATION:
; APPLICANT: Bauer, David
; APPLICANT: Collmer, Alan
; TITLE OF INVENTION: Hypersensitive Response Elicitor
; TITLE OF INVENTION: From
; TITLE OF INVENTION: Erwinia Chrysanthemi
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,358
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/840
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1304
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-484-358-2

Query Match 32.0%; Score 718.5; DB 2; Length 338;
Best Local Similarity 42.8%; Pred. No. 8.8e-56;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;
QY 47 MQISI-GGAGGNNGLLGTSRQNALGG-NSA---LGLGGNQNDVTYNOLAGLLTGMMMM 101
Db 1 MQITKAHIGDGLGVSLGAQ--GLKGLNSAASSLGSSVDKLSSTIDKLTSAITSM--- 55
QY 102 SMGGGGLMGGGLGNGLGGSGGLGEGLSNALNDMLGSLNTLGSKGNNTTSTTNS 161
Db 56 -----FGALAQGLGAS-SKGLGMSNQLGQSFGN-----GAQGASNLLSVPK- 96
QY 162 PLDQALGINSTSQNDSTSGTDTSDSDPMQQLKMFSEIMQSLFG-----DGDG 213
Db 97 -----SGDALS-----KMFDAKLDLLGHDVTVKLTNQSNQ 128
QY 214 TQSSSGGKQPTGEQNAKKVYTDALSLGMLNGLSOLLGNGLGGGGGAGTGLDGS 273
Db 129 LANSMLNASQMTQGNMNAFSGVNNALSSILGNGLSGM-----SGFSQPS 174
QY 274 LGKGLQNLSPVDYQQLGNNAVGTGIGMKAGIQALNDIGTHRSSTFRSVNKGDRMAKE 333
Db 175 LGAGLQGLSGAGAFNLGNAGVGVQNALSALSNVTHVDGNRRHFVDKDRGMAKE 234
QY 334 IGFMDQYPEVFGKPYQKPGGQEVKTTDDKSWAKLSKPDGDMTPASMEQFNKAGMIK 393
Db 235 IGFMDQYPEIFGKPEYQKDGWSSPKTDDKSWAKLSKPDGDMTPASMEQFNKAGMIK 294

QY 394 RPWAGDTGNGLQARGAGGSSGLGIDAMAGDAINNMALGKLGAA 437
Db 295 SAVAGDTGNTNLNLRCGAGGASLGIDAADVGDKIANNMGLKLANA 338
RESULT 12
US-08-819-539-1
; Sequence 1, Application US/08819539
; Patent No. 5859324
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: Hypersensitive Response
; TITLE OF INVENTION: Induced Resistance In Plants
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/819,539
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/475,775
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 14603/10050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-819-539-1

Query Match 32.0%; Score 718.5; DB 2; Length 338;
Best Local Similarity 42.8%; Pred. No. 8.8e-56;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;
QY 47 MQISI-GGAGGNNGLLGTSRQNALGG-NSA---LGLGGNQNDVTYNOLAGLLTGMMMM 101
Db 1 MQITKAHIGDGLGVSLGAQ--GLKGLNSAASSLGSSVDKLSSTIDKLTSAITSM--- 55
QY 102 SMGGGGLMGGGLGNGLGGSGGLGEGLSNALNDMLGSLNTLGSKGNNTTSTTNS 161
Db 56 -----FGALAQGLGAS-SKGLGMSNQLGQSFGN-----GAQGASNLLSVPK- 96
QY 162 PLDQALGINSTSQNDSTSGTDTSDSDPMQQLKMFSEIMQSLFG-----DGDG 213
Db 97 -----SGDALS-----KMFDAKLDLLGHDVTVKLTNQSNQ 128
QY 214 TQSSSGGKQPTGEQNAKKVYTDALSLGMLNGLSOLLGNGLGGGGGAGTGLDGS 273
Db 129 LANSMLNASQMTQGNMNAFSGVNNALSSILGNGLSGM-----SGFSQPS 174
QY 274 LGKGLQNLSPVDYQQLGNNAVGTGIGMKAGIQALNDIGTHRSSTFRSVNKGDRMAKE 333
Db 175 LGAGLQGLSGAGAFNLGNAGVGVQNALSALSNVTHVDGNRRHFVDKDRGMAKE 234

```

OY 274 LGCKLONLSPVDYOOGLNAGVTGICMKAGIOALNDIGTHRHSSSTFVSFNKGDRAWAKE 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 175 LGAGGLQGLSGAGAFNLGNAIGNGVGQNAALGALSNSVSTHVDGNHRHFVDKEDRCMAKE 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 334 IGQFMDQYVEFGKPYQYQKPGQEVETKDDKSWAKALSKPDDDDGMTPASMEQFNKAKGMK 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 235 IGQFMDQYPEIFGKPEYQKDGWSSPKTDDKSWAKALSKPDDDDGMTGASMDKFRQANGMIK 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 394 RPMAGDTGNLQARGAGGSSLGIDAMAGDAINNALGKLGAA 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 295 SAVACDRGNTNLNRGAGGASLGIDAAVVGDKIANMSLGKLANA 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-09-118-959-2
; Sequence 2, Application US/09118959
; Patent No. 6001959
; GENERAL INFORMATION:
; APPLICANT: Bauer, David
; APPLICANT: Collmer, Alan
; TITLE OF INVENTION: Hypersensitive Response Elicitor From
; TITLE OF INVENTION: Erwinia Chrysanthemi
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/118,959
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/840
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1304
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-118-959-2

```

```

Query Match      32.0%; Score 718.5; DB 3; Length 338;
Best Local Similarity 42.8%; Pred. No. 8.8e-56;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;

47 MQTISI-QCAGGNNCLLTSTPQNAAGLG-NSA---LGLGGGNQNDTWNQAGLITGMMMM 101
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 MQITIKAHIGDGLGVSGLGAQ--GLKGLNSAASSLESSVDKLSSTIDKLTSLTSMW--- 55
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

102 SMHGGGGLMGGGGLGNGLGSGGLGEGLSNALNDMLGGSLNTLGGKGNNTSTWNS 161
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
56 -----EGGALAAOGLGAS-SKGLGMSNOLGOSTGN-----GAQCA SNLLSVPK- 96
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

162 PLQALGINSTQNDDSTSGTSDTSDSSDPMQOLLKMFSEIMQSLFG-----DQDQG 213
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
97 -----SGGDALS-----RMFKALDDLLGHDTVTKTNQNSQ 128
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY 214 TQSSSGGKQPTGEQONAYKKGVTDALSLMGNLSOLLNGGLGGGGNAGTGLDGSS 273
Db 129 LANSMLNASQMTQGNMNAFGSGVNNALSSILNGLGQSM-----SGFSQPS 174
QY 274 LGGKGLQNLSPVDYQOLGNVAGTGTGIGMKAGTQALNDIGTTHRSSTRSFVNKGDRAKAKE 333
Db 175 LGAGGLQSLGAGAFNOLGNAIGMGVGNALSNVSTHYDGNRRHFVDKEDRCMAKAKE 234
QY 334 IGQFMDQYPEVFGKPYQKGPQEVKTDKSWAKALSKPDDDMTPASMEQFNKAKGMK 393
Db 235 IGQFMDQYPEIFGKPYQKGDWSSPKTDDKSWAKALSKPDDDMTPASMEQFNKAKGMK 294
QY 394 RPMAGDTGNLQARGAGSSGLGIDAMMAGDAINNMALGKLGAA 437
Db 295 SAVAGDTGNTNLNRGAGGASLGIDAAVVGDKIANMSLGKLANA 338

RESULT 15

US-08-984-207-1
; Sequence 1, Application US/08984207
; Patent No. 6235974
; GENERAL INFORMATION:
; APPLICANT: Qiu, Dwen
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
; TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/984,207
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,230
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1201
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-984-207-1

Query Match 32.0%; Score 718.5; DB 4; Length 338;
Best Local Similarity 42.8%; Pred. No. 8.8e-56;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;
QY 47 MQIST-GGAGGNLIGTSRQNAIGLG-NSA--LGLGGGNQNDVNLQAGLLTGMMMM 101
Db 1 MQITKAHIGGDLGVSGLAGQ--GLKGLNSAASSIGSSVDKLSSTIDKLTSLTSM--- 55
QY 102 SMGGGGGLMGGGLGGLGNGGLGGGLGEGLSNALNDMLGGSNLTLGSKGGNNTTSTTNS 161

Db 56 -----FGGALAOQLGAS-SKGLGMSNOLGOSFGN-----CAQGASNLLSVPK- 96
QY 162 PLDOALGINSTSONDDSTSGTDTSDSSDPMQQLLKMFEIMQSLFG-----DQDQG 213
Db 97 -----SGGDALS-----KMFDAKALDDLLGHDTVTKLTNQSNQ 128
QY 214 TQSSSGGKQPTGEQONAYKKGVTDALSLMGNLSOLLNGGLGGGGNAGTGLDGSS 273
Db 129 LANSMLNASQMTQGNMNAFGSGVNNALSSILNGLGQSM-----SGFSQPS 174
QY 274 LGGKGLQNLSPVDYQOLGNVAGTGTGIGMKAGTQALNDIGTTHRSSTRSFVNKGDRAKAKE 333
Db 175 LGAGGLQSLGAGAFNOLGNAIGMGVGNALSNVSTHYDGNRRHFVDKEDRCMAKAKE 234
QY 334 IGQFMDQYPEVFGKPYQKGPQEVKTDKSWAKALSKPDDDMTPASMEQFNKAKGMK 393
Db 235 IGQFMDQYPEIFGKPYQKGDWSSPKTDDKSWAKALSKPDDDMTPASMEQFNKAKGMK 294
QY 394 RPMAGDTGNLQARGAGSSGLGIDAMMAGDAINNMALGKLGAA 437
Db 295 SAVAGDTGNTNLNRGAGGASLGIDAAVVGDKIANMSLGKLANA 338

Search completed: January 22, 2003, 16:26:34
Job time : 19 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:26:07 ; Search time 40 Seconds
(without alignments)
1455.763 Million cell updates/sec

Title: 693_11_3

Perfect score: 2248

Sequence: 1 MGFFLFQMPSEFLVSTLLL.....DAMMAGDAINNMAKLGAA 437

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

```

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2079	92.5	403	19 AAW75863	Erwinia amylovora
2	2079	92.5	403	19 AAW62455	Erwinia amylovora
3	2079	92.5	403	19 AAW61114	Hypersensitive res
4	2079	92.5	403	20 AAW87639	A hypersensitive r
5	2079	92.5	403	21 AAY71093	Erwinia amylovora
6	2079	92.5	403	21 AAY84854	A hypersensitive r
7	2079	92.5	403	22 AAE06710	Erwinia amylovora
8	2079	92.5	403	23 ABB09225	Erwinia amylovora
9	2079	92.5	403	23 AAE18295	Erwinia amylovora
10	2079	92.5	403	23 AAE16447	E. amylovora hyper

11	1928	85.8	385	18	AAW06598	Hypersensitive res
12	1913	85.1	385	15	AAW45751	Erwinia amylovora
13	718.5	32.0	338	18	AAW06597	Hypersensitive res
14	718.5	32.0	338	20	AAW87638	A hypersensitive r
15	718.5	32.0	338	20	AAW82407	E. chrysanthemi Hr
16	718.5	32.0	338	21	AAW71092	Erwinia chrysanth
17	718.5	32.0	338	21	AAW84853	A hypersensitive r
18	718.5	32.0	338	21	AAW55801	E. chrysanthemi hy
19	718.5	32.0	338	22	AAE06709	Erwinia chrysanth
20	718.5	32.0	338	23	ABB09224	Erwinia chrysanth
21	718.5	32.0	338	23	AAE18294	Erwinia chrysanth
22	718.5	32.0	338	23	AAE16446	E. chrysanthemi hy
23	718.5	32.0	340	19	AAW75862	Erwinia chrysanth
24	718.5	32.0	340	19	AAW61113	Hypersensitive res
25	713.5	31.7	340	19	AAW62454	Erwinia chrysanth
26	226.5	10.1	898	18	AAW31853	Mycobacterium tube
27	224.5	10.0	272	22	ABB70198	Drosophila melanog
28	215.5	9.6	718	12	AAW14308	N. clavipes draglin
29	211.5	9.4	651	20	AAV40097	Spider silk protei
30	211.5	9.4	651	23	AAU11781	Spider natural sil
31	211.5	9.4	718	19	AAW53346	Nephila clavipes s
32	211.5	9.4	718	21	AAV59070	N. clavipes spider
33	206	9.2	604	16	AAW99057	Spider dragline va
34	204	9.1	528	22	ABB82611	Spider recombinant
35	203.5	9.1	318	21	ABB81229	Mycobacterium tube
36	202	9.0	630	23	AAW50042	N. clavipes spidro
37	202	9.0	676	23	AAW50047	N. clavipes spidro
38	202	9.0	989	23	AAW50038	N. clavipes spidro
39	202	9.0	1255	23	AAW50037	N. clavipes spidro
40	202	9.0	1880	23	AAW50039	N. clavipes spidro
41	202	9.0	2280	22	ABB61650	Drosophila melanog
42	201.5	9.0	5002	22	ABB63723	Mycobacterium spec
43	200	8.9	388	20	AAW04999	Arabidopsis thalia
44	199.5	8.9	201	21	AAG31750	New DNA sequence i
45	198.5	8.8	738	19	AAW56163	

ALIGNMENTS

RESULT 1
AAW75863
ID AAW75863 standard; Protein; 403 AA.
XX AAW75863:
XX
XX 07-DEC-1998 (first entry)
DT
DE Erwinia amylovora hypersensitive response elicitor (HRE).
DE
XX
KW Hypersensitive response elicitor; HRE; insect resistance.
KW Biological control; transgenic plant.
XX
OS Erwinia amylovora.
XX
PN WO9837752-A1.
XX
PD 03-SEP-1998.
XX
PF 26-FEB-1998; 98WO-US03604.
XX
PR 28-FEB-1997; 97US-0039226.
XX
(CORR) CORNELL RES FOUND INC.
XX
PI Wei Z, Zitter TA;
XX
DR WPI; 1998-495374/42.
DR N-PSDB; AAV54607.
XX
XX use of hypersensitive response elicitor polypeptide - for
PT application to plants or seeds or transgenic plants or seeds for the
PT control of insects.

XX DT 26-OCT-1998 (first entry)
XX DE Hypersensitive response elicitor protein (39 kDa).
XX KW Hypersensitive response elicitor; transgenic plant; seed;
XX KW pathogen resistance; disease resistance; crop protection.
OS Erwinia amylovora.
XX WO9824297-A1.
PN 11-JUN-1998.
PD 04-DEC-1997; 97WO-US22629.
PF 05-DEC-1996; 96US-0033230.
PR (CORR) CORNELL RES FOUND INC.
XX Beer SV, Qiu D, Wei Z;
PI WPI; 1998-332931/29.
XX N-PSDB; AAV36428.
DR Imparting pathogen resistance to plants - by applying a
PT hypersensitive response elicitor polypeptide to seeds
PS Disclosure; Page 18-20; 85pp; English.
XX This is the 39 kDa hypersensitive elicitor (HRE) protein of
CC Erwinia amylovora. It is heat stable at 100 degC for at least 10
CC min, has a PI of approximately 4.3, and contains substantially no
CC cysteine. The invention relates to methods of imparting
CC hypersensitive response induced resistance to plants by treatment
CC of seeds. Isolated HRE proteins can be applied to seeds as a means
CC of imparting pathogen resistance to plants grown from the seeds.
CC Alternatively, bacteria containing the gene encoding the HRE can be
CC applied to the plant seeds, or transgenic plant seeds containing a
CC DNA molecule encoding an HRE polypeptide or protein are used. HRE
CC pseudomonas syringae, Pseudomonas solanacearum, Xanthomonas
CC campestris pv. glycines and Xanthomonas campestris pelargonii (see
CC AA61113-18) are provided. The methods can impart pathogen
CC resistance without using agents which are harmful to the
CC environment or pathogenic to the plant seed being treated, or to
CC adjacent plants. E. amylovora causes disease in apple or pear
CC but not tomato. However, it elicits a hypersensitive response in
CC tomato. Thus, E. amylovora can be applied to tomato seeds to
CC impart pathogen resistance without causing diseases in plants of
CC that species.
XX SQ Sequence 403 AA;
* Query Match 92.5%; Score 2079; DB 19; Length 403;
Best Local Similarity 100.0%; Pred. No. 3.5e-155;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 MSLNTSLGCASTMQISIGGAGGNNLLGTSRQNALGNSALGLGGGNQNTVNLGAGLL 94
Db 1 MSLNTSLGCASTMQISIGGAGGNNLLGTSRQNALGNSALGLGGGNQNTVNLGAGLL 60
QY 95 TGMNMMNMGVGGGLMGGGLGGGLGNGLGGSGGLGEGLSNALNDMLGGSNLTLGSKGNN 154
Db 61 TGMNMMNMGVGGGLMGGGLGGGLGNGLGGSGGLGEGLSNALNDMLGGSNLTLGSKGNN 120
QY 155 TTSTNTPLDQALGINSTQNDSTSGTDSFSDSDPMQQLKWFSEIMQSLFGDGQDGT 214
Db 121 TTSTNTPLDQALGINSTQNDSTSGTDSFSDSDPMQQLKWFSEIMQSLFGDGQDGT 180
QY 215 QGSSSGKQPTGEGQNAKKGVTDALSLMGNGLSQLLNGNLGGGCGGNAGTGLDGSSL 274
Db 181 QGSSSGKQPTGEGQNAKKGVTDALSLMGNGLSQLLNGNLGGGCGGNAGTGLDGSSL 240

QY 275 GKGGLQNLGSPVDYQOQLGNAVGTGIGMKAGIQALNDIGTHRHSSSTRSFVNKGDRAMAKEI 334
Db 241 GKGGLQNLGSPVDYQOQLGNAVGTGIGMKAGIQALNDIGTHRHSSSTRSFVNKGDRAMAKEI 300
QY 335 GGFMDQYPEVFGKPOYKGPQGVKTDKSKWAKALSKPDDGTMTPASMEQFNKAGMIKR 394
Db 301 GGFMDQYPEVFGKPOYKGPQGVKTDKSKWAKALSKPDDGTMTPASMEQFNKAGMIKR 360
QY 395 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437
Db 361 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 403

RESULT 4
AAW87639
ID AAW87639 standard; Protein; 403 AA.
XX AC AAW87639;
XX DT 09-MAR-1999 (first entry)
XX DE A hypersensitive response elicitor protein.
KW Hypersensitive response elicitor protein; hairpin protein;
KW disease resistance; seed quality; insect control; corn borer;
KW Lepidoptera larvae; transgenic plant.
XX OS Erwinia amylovora.
XX PN WO9854214-A2.
XX PD 03-DEC-1998.
XX PF 28-MAY-1998; 98WO-US10874.
XX PR 30-MAY-1997; 97US-0048109.
XX PA (CORR) CORNELL RES FOUND INC.
XX PA (EDEN-) EDEN BIOSCIENCE CORP.
XX PI Beer SV, Laby RJ, Wei Z;
XX WPI; 1999-070210/06.
XX N-PSDB; AAV83989.
New fragments of an Erwinia hypersensitive response elicitor protein
and related DNA - used to impart disease resistance to plants, to
increase their growth and to control insects
Claim 4; Page 10-11; 94pp; English.

XX The present sequence represents a hypersensitive response elicitor
CC protein (also called hairpin protein) that is able to elicit a
CC hypersensitive response in plants. The specification also describes
CC hypersensitive response elicitors from other pathogenic organisms.
CC The protein, in non-infectious form, is applied to plants to impart
CC disease resistance (to a wide range of viral, bacterial and fungal
CC pathogens), to improve growth (yield, quantity and quality of seeds,
CC to provide earlier germination etc.) and to control insects (e.g. corn
CC borers, Lepidoptera larvae etc.) The same results are provided by
CC transgenic plants expressing the protein.

XX SQ Sequence 403 AA;
Query Match 92.5%; Score 2079; DB 20; Length 403;
Best Local Similarity 100.0%; Pred. No. 3.5e-155;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 MSLNTSLGCASTMQISIGGAGGNNLLGTSRQNALGNSALGLGGGNQNTVNLGAGLL 94
Db 1 MSLNTSLGCASTMQISIGGAGGNNLLGTSRQNALGNSALGLGGGNQNTVNLGAGLL 60

XX WPI: 2000-303745/26.
DR N-PSDB; AAA14938.
XX
PT Hypersensitive response elicitor polypeptides useful for imparting
PT enhanced growth, disease resistance and insect resistance to plants,
PT especially vegetables and ornamental flowers -
XX
XX Claim 4; Page 8-10; 100pp; English.
XX
XX The present sequence represents a hypersensitive response elicitor
CC polypeptide. The specification describes hypersensitive response
CC elicitor polypeptide fragments, which do not elicit a hypersensitive
CC response. Instead, the proteins impart disease resistance to plants,
CC enhance plant growth, and/or control insects. The polypeptide
CC fragments may be used to these properties to plants. The plants which
CC may be treated in this way include vegetables, crops and ornamental
CC plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,
CC peanut, corn, potato, sweet potato, bean, pea, chichory, lettuce,
CC endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,
CC broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,
CC carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,
CC citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,
CC petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.
XX
XX Sequence 403 AA;
SQ
Query Match 92.5%; Score 2079; DB 21; Length 403;
Best Local Similarity 100.0%; Pred. No. 3.5e-155;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 MSLNTSGLGASTMQISIGGAGGNNLLGTSRQNALGGSALGLGGGNQNDTVNQLAGLL 94
Db 1 MSLNTSGLGASTMQISIGGAGGNNLLGTSRQNALGGSALGLGGGNQNDTVNQLAGLL 60
QY 95 TGMMMMSMMGGGLMGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 154
Db 61 TGMMMMSMMGGGLMGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 120
QY 155 TTSTTNSPLDQALGINSTSONDDSTSGTSDSSDPMQOLLKMFSEIMQSLFGDGQDGT 214
Db 121 TTSTTNSPLDQALGINSTSONDDSTSGTSDSSDPMQOLLKMFSEIMQSLFGDGQDGT 180
QY 215 QGSSSGGKQPTGEONAYKGVTDALSGLMGNLSQLLGGNGGGGGGAGTGLDGSLL 274
Db 181 QGSSSGGKQPTGEONAYKGVTDALSGLMGNLSQLLGGNGGGGGGAGTGLDGSLL 240
QY 275 GKGGLQNLGSPVDYQQLGNVAVGTGIGMKAGIQALNDIGTHRHSSSTRSFVNKGDRAAKEI 334
Db 241 GKGGLQNLGSPVDYQQLGNVAVGTGIGMKAGIQALNDIGTHRHSSSTRSFVNKGDRAAKEI 300
QY 335 GQFMDQYPEVFGKPOYQKGPQGEVKTDDKSWAKALSKPDDDGMTFASMEQFNKAKGMIKR 394
Db 301 GQFMDQYPEVFGKPOYQKGPQGEVKTDDKSWAKALSKPDDDGMTFASMEQFNKAKGMIKR 360
QY 395 PMAGDTGNGNLQARGAGGSSLGIDAMMAGDAINNMALGKLGA 437
Db 361 PMAGDTGNGNLQARGAGGSSLGIDAMMAGDAINNMALGKLGA 403
RESULT 7
AAE06710
ID AAE06710 standard; Protein; 403 AA.
XX
AC AAE06710;
XX
DT 16-OCT-2001 (first entry)
XX
DE Erwinia amylovora hypersensitive response elicitor protein.
XX
KW Hypersensitive response elicitor; oomycete; transgenic plant; infection;
KW gene therapy; crop loss; antifungal.

XX Erwinia amylovora.
XX WO200155347-A1.
XX
PD 02-AUG-2001.
XX
XX 26-JAN-2001; 2001WO-US02579.
XX
XX 26-JAN-2000; 2000US-0178565.
XX
XX (CORR) CORNELL RES FOUND INC.
FA
PI Beer SV, Bauer DW;
XX
XX WPI: 2001-488791/53.
DR N-PSDB; AAD12806.
XX
XX New chimeric gene, useful for controlling plant-pathogenic fungi and
PT producing oomycete-resistant transgenic plants, comprises first DNA
PT encoding hypersensitive response elicitor, promoter and regulatory
PT region -
XX
PS Claim 9; Page 13-14; 72pp; English.
XX
XX The invention relates to a chimeric gene that includes a first DNA
CC molecule encoding a hypersensitive response elicitor protein or
CC polypeptide, promoter operably linked 5' to the first DNA molecule
CC to induce transcription of the first DNA molecule in response to
CC activation of the promoter by an oomycete and a 3' regulatory region
CC operably linked to the first DNA molecule. The invention also relates
CC to a transgenic plant resistant to disease resulting from oomycete
CC infection, the transgenic plant including the chimeric gene, wherein
CC the promoter induces transcription of the first DNA molecule in
CC response to infection of the plant by an oomycete. The chimeric gene
CC is used in gene therapy. The chimeric gene is useful as an effective
CC and safe means of controlling plant-pathogenic fungi, particularly
CC oomycetes, which are responsible for major crop loss and is also useful
CC for producing transgenic plants of the invention. The present sequence
CC is Erwinia amylovora hypersensitive response elicitor protein.
XX
SQ Sequence 403 AA;
Query Match 92.5%; Score 2079; DB 22; Length 403;
Best Local Similarity 100.0%; Pred. No. 3.5e-155;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 MSLNTSGLGASTMQISIGGAGGNNLLGTSRQNALGGSALGLGGGNQNDTVNQLAGLL 94
Db 1 MSLNTSGLGASTMQISIGGAGGNNLLGTSRQNALGGSALGLGGGNQNDTVNQLAGLL 60
QY 95 TGMMMMSMMGGGLMGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 154
Db 61 TGMMMMSMMGGGLMGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGGGLGG 120
QY 155 TTSTTNSPLDQALGINSTSONDDSTSGTSDSSDPMQOLLKMFSEIMQSLFGDGQDGT 214
Db 121 TTSTTNSPLDQALGINSTSONDDSTSGTSDSSDPMQOLLKMFSEIMQSLFGDGQDGT 180
QY 215 QGSSSGGKQPTGEONAYKGVTDALSGLMGNLSQLLGGNGGGGGGAGTGLDGSLL 274
Db 181 QGSSSGGKQPTGEONAYKGVTDALSGLMGNLSQLLGGNGGGGGGAGTGLDGSLL 240
QY 275 GKGGLQNLGSPVDYQQLGNVAVGTGIGMKAGIQALNDIGTHRHSSSTRSFVNKGDRAAKEI 334
Db 241 GKGGLQNLGSPVDYQQLGNVAVGTGIGMKAGIQALNDIGTHRHSSSTRSFVNKGDRAAKEI 300
QY 335 GQFMDQYPEVFGKPOYQKGPQGEVKTDDKSWAKALSKPDDDGMTFASMEQFNKAKGMIKR 394
Db 301 GQFMDQYPEVFGKPOYQKGPQGEVKTDDKSWAKALSKPDDDGMTFASMEQFNKAKGMIKR 360
QY 395 PMAGDTGNGNLQARGAGGSSLGIDAMMAGDAINNMALGKLGA 437
Db 361 PMAGDTGNGNLQARGAGGSSLGIDAMMAGDAINNMALGKLGA 403

Db 361 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMAKGKLGAA 403

RESULT 8

ABB09225

ID ABB09225 standard; Protein; 403 AA.

XX ABB09225;

AC ABB09225;

XX 08-JUL-2002 (first entry)

DE Erwinia amylovora hypersensitive response elicitor SRO ID NO:3.

XX Erwinia amylovora hypersensitive response elicitor; fruit; vegetable; plant; desiccation;

XX postharvest disease.

KW Hypersensitive response elicitor; fruit; vegetable; plant; desiccation;

XX postharvest disease.

OS Erwinia amylovora.

XX Erwinia amylovora.

XX WO200180639-A2.

PN 01-NOV-2001.

XX 17-APR-2001; 2001WO-US12468.

PF 19-APR-2000; 2000US-198359P.

XX (EDEN-) EDEN BIOSCIENCE CORP.

PA Wei Z, Qiu D, Remick D;

PI WPI; 2002-041357/05.

XX N-PSDB; ABL51710.

DR Inhibiting post harvest disease (caused by Penicillium, Botrytis,

XX Phytophthora, or Erwinia) or desiccation and enhancing the longevity in

PT a fruits or vegetables, using hypersensitive response elicitor proteins

PT or nucleic acids -

XX Example; Page 9-10; 72pp; English.

PS The present invention describes methods for inhibiting post harvest

XX disease or desiccation and enhancing the longevity in a fruits or

XX vegetables, using hypersensitive response elicitor proteins or

CC polypeptides or nucleic acids (I) derived from pathogens (e.g. Erwinia

CC amylovora, E. stewartii, E. chrysanthemi, E. carotovora, Xanthomonas,

CC Pseudomonas syringae, P. solanacearum, Phytophthora, and Clavibacter).

CC (I) has bactericidal activity, and can be used in gene therapy. The

CC method can be used for inhibiting post harvest disease (caused by

CC Penicillium, Botrytis, Phytophthora, or Erwinia) or desiccation and

CC enhancing the longevity in a fruits or vegetables. The method enables

CC growers, warehouse packers, shippers and suppliers to process, handle

CC and store fruit and vegetables with reduced losses caused by post harvest

CC disease and desiccation, therefore reducing costs to the consumer and

CC improving quality. The present sequence represents a hypersensitive

CC response elicitor protein given in the exemplification of the present

XX invention.

XX Sequence 403 AA;

Query Match 92.5%; Score 2079; DB 23; Length 403;

Best Local Similarity 100.0%; Pred. No. 3.5e-155;

Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 MSLNTSGLAGSTMTQISTGGAGNGLLGTSRQNALGNSALGLGGNQNDVTYNLAGLL 94

DB 1 MSLNTSGLAGSTMTQISTGGAGNGLLGTSRQNALGNSALGLGGNQNDVTYNLAGLL 60

OY 95 TGMAMMMVMCCGGLMCGGLCGGLCGGLCGGLCGGLCGGLCGGLCGGLCGGLCGGLCGGL 154

DB 61 TGMAMMMVMCCGGLMCGGLCGGLCGGLCGGLCGGLCGGLCGGLCGGLCGGLCGGL 120

OY 155 TTSTTNSPLDQALGINSTQNDSTSGTSDSSDPMOQLLKMFSEIMOSLFGDQDGT 214

Db 121 TTSTTNSPLDQALGINSTQNDSTSGTSDSSDPMOQLLKMFSEIMOSLFGDQDGT 180

OY 215 QGSSSGGKOPTGEQNAKKGYTDALSGLMGNGLSOLLNGGLGGGQGGNAGTGLDGSSL 274

DB 181 QGSSSGGKOPTGEQNAKKGYTDALSGLMGNGLSOLLNGGLGGGQGGNAGTGLDGSSL 240

OY 275 GKGGLONLSGPVYDQOLGNVAGTIGMKKAGIOALNDIGTHRSSSTRSFYNKGDRAKKEI 334

DB 241 GKGGLONLSGPVYDQOLGNVAGTIGMKKAGIOALNDIGTHRSSSTRSFYNKGDRAKKEI 300

OY 335 GGFMDQYPEVFGKPOYKOGPQGVKTDKSKAKLSKDDDDGCTPDASMEQFNKAKGMIKR 304

DB 301 GGFMDQYPEVFGKPOYKOGPQGVKTDKSKAKLSKDDDDGCTPDASMEQFNKAKGMIKR 360

OY 395 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMAKGKLGAA 437

DB 361 PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMAKGKLGAA 403

RESULT 9

AAE18295

ID AAE18295 standard; Protein; 403 AA.

XX AAE18295;

AC AAE18295;

XX 07-MAY-2002 (first entry)

DE Erwinia amylovora hypersensitive response elicitor (HRE) #1.

XX Hypersensitive response elicitor; HRE; transgenic plant; plant growth;

KW stress tolerance; disease tolerance; modified flower colour;

KW insect resistance; herbicide resistance; male sterility.

XX Erwinia amylovora.

OS Erwinia amylovora.

XX WO2000195724-A2.

PN 20-DEC-2001.

XX 13-JUN-2001; 2001WO-US18955.

PF 15-JUN-2000; 2000US-211585P.

XX (EDEN-) EDEN BIOSCIENCE CORP.

PA Wei Z, Derocher J;

PI WPI; 2002-130707/17.

XX N-PSDB; AAD29124.

DR Improving effectiveness of transgenic plants by topical application of

XX a hypersensitive response elicitor protein to the transgenic plant or

XX by incorporating into the plant a transgene encoding the protein -

PS Disclosure; Page 10-11; 86pp; English.

XX The invention relates to methods of improving the effectiveness of

CC transgenic plants which involves either topical application of a

CC hypersensitive response elicitor (HRE) protein to the transgenic plant

CC or incorporating into the transgenic plant a transgene encoding HRE.

CC HRE sequence is used for improving the effectiveness of transgenic

CC plants by maximising the benefit of transgenic traits associated with

CC a deleterious effect on growth, stress tolerance, disease or insect

CC resistance, enhanced growth, herbicide resistance, male sterility,

CC modified flower colour and biochemically modified plant product in

CC the transgenic plants or overcoming the deleterious effects. The

XX present sequence is Erwinia amylovora HRE protein.

XX Sequence 403 AA;

Query Match 92.5%; Score 2079; DB 23; Length 403;

Best Local Similarity 100.0%; Pred. No. 3.5e-155;

Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	35	MSLNTSGLGASTMQISIGGAGGNLLGTSRONAGLGGNSALGLGGGNQNDTVNQLAGLL	94
Db	1	MSLNTSGLGASTMQISIGGAGGNLLGTSRONAGLGGNSALGLGGGNQNDTVNQLAGLL	60
QY	95	TGMMMMMSMGGGLMGGGLGGLGNGTGGGGLGEGLSNALNDMLGGSNTLGSKGGNN	154
Db	61	TGMMMMMSMGGGLMGGGLGGLGNGTGGGGLGEGLSNALNDMLGGSNTLGSKGGNN	120
QY	155	TTSTTNSPLDQALGINSTSQNDSTSGTSDTSDSDPMQQLKMFSEIMQSLFGDGQDGT	214
Db	121	TTSTTNSPLDQALGINSTSQNDSTSGTSDTSDSDPMQQLKMFSEIMQSLFGDGQDGT	180
QY	215	QGSSSGGKQPTGEQNAKKGVTDALSLGMNGLSQLLGGNGGLGGGQGNAGTGLDGSSL	274
Db	181	QGSSSGGKQPTGEQNAKKGVTDALSLGMNGLSQLLGGNGGLGGGQGNAGTGLDGSSL	240
QY	275	GGKGLQNLSPVDYQQLGNNAVGTGIGMKAGIQALNDIGTHRHSSTRSFVNKGDRAMAKEI	334
Db	241	GGKGLQNLSPVDYQQLGNNAVGTGIGMKAGIQALNDIGTHRHSSTRSFVNKGDRAMAKEI	300
QY	335	QGFMDQYPEVFGKPOYQKPGQEVKTDKSWAKALSKPDDGWTGPASMEQFNKAKGMIKR	394
Db	301	QGFMDQYPEVFGKPOYQKPGQEVKTDKSWAKALSKPDDGWTGPASMEQFNKAKGMIKR	360
QY	395	PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437	
Db	361	PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 403	
DR	AAE16447	AAE16447 standard; Protein; 403 AA.	
XX	AAE16447;		
AC	AAE16447;		
DT	09-APR-2002	(first entry)	
XX	E. amylovora hypersensitive response elicitor protein #1.		
DE	Hypersensitive response elicitor protein; plant growth; fruit coloration;		
KW	disease resistance; stress resistance; phytoalexin; insect infection;		
KW	plant maturation.		
XX	Erwinia amylovora.		
OS	Location/Qualifiers		
FH	Key		
FT	Domain		
FT	/label= Hypersensitive_response_eliciting_domain_1		
FT	32..74		
FT	Region		
FT	/label= Acidic_unit		
FT	57..74		
FT	Region		
FT	/label= Alpha_helix		
FT	130..180		
FT	Domain		
FT	/label= Hypersensitive_response_eliciting_domain_2		
FT	130..157		
FT	Region		
FT	/label= Acidic_unit		
FT	157..180		
FT	Region		
FT	/label= Alpha_helix		
XX	WO200198501-A2.		
PN	27-DEC-2001.		
XX	12-JUN-2001; 2001WO-US18820.		
XX	16-JUN-2000; 2000US-212211P.		
XX	(EDEN-) EDEN BIOSCIENCE CORP.		
PA	Fan H, Wei Z;		
PI	WPI; 2002-122282/16.		
XX			
DR			

DR	N-PSDB; AAD27015.		
XX	New hypersensitive response elicitor proteins comprising spaced apart		
PT	domains having an acidic portion linked to an alpha-helix, useful for		
PT	impacting disease or stress resistance, controlling insects or		
PT	enhancing plant growth -		
XX	Disclosure; Page 9-10; 99pp; English.		
PS	The patent discloses hypersensitive response elicitor proteins and		
XX	nucleotides encoding such proteins. Hypersensitive response elicitor		
CC	proteins comprise an isolated pair or more of spaced apart domains,		
CC	each comprising an acidic portion linked to an alpha-helix and capable		
CC	of eliciting a hypersensitive response in plants. Sequences of the		
CC	invention are used to impart disease resistance to plants, to enhance		
CC	plant growth, to control insects and/or to impart stress resistance		
CC	to plants which includes resistance to environmental stresses such as		
CC	climate, air pollution, chemical and nutritional stress. The method of		
CC	impacting disease resistance has the potential for treating previously		
CC	untreatable diseases, treating diseases systemically and avoiding the		
CC	use of infectious agents or environmentally harmful materials. Hyper-		
CC	sensitive response elicitor sequences are used to enhance plant growth		
CC	which encompasses greater yield, increased in quantity of seeds produced,		
CC	percentage of seeds germinated, plant size and biomass, bigger fruits,		
CC	earlier fruit coloration and plant maturation. They are also used for		
CC	insect control which encompasses preventing direct insect damage to		
CC	plants by feeding injury, interfering with insect larval feeding on the		
CC	plants, preventing insects from colonising host plants and releasing		
CC	phytochemicals. Sequences of the invention also prevent subsequent disease		
CC	damage to plants resulting from insect infection. The present sequence		
CC	is Erwinia amylovora hypersensitive response elicitor protein.		
XX	Sequence 403 AA;		
SQ	Query Match 92.5%; Score 2079; DB 23; Length 403;		
	Best Local Similarity 100.0%; Pred. No. 3.5e-155;		
	Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	35	MSLNTSGLGASTMQISIGGAGGNLLGTSRONAGLGGNSALGLGGGNQNDTVNQLAGLL	94
Db	1	MSLNTSGLGASTMQISIGGAGGNLLGTSRONAGLGGNSALGLGGGNQNDTVNQLAGLL	60
QY	95	TGMMMMMSMGGGLMGGGLGGLGNGTGGGGLGEGLSNALNDMLGGSNTLGSKGGNN	154
Db	61	TGMMMMMSMGGGLMGGGLGGLGNGTGGGGLGEGLSNALNDMLGGSNTLGSKGGNN	120
QY	155	TTSTTNSPLDQALGINSTSQNDSTSGTSDTSDSDPMQQLKMFSEIMQSLFGDGQDGT	214
Db	121	TTSTTNSPLDQALGINSTSQNDSTSGTSDTSDSDPMQQLKMFSEIMQSLFGDGQDGT	180
QY	215	QGSSSGGKQPTGEQNAKKGVTDALSLGMNGLSQLLGGNGGLGGGQGNAGTGLDGSSL	274
Db	181	QGSSSGGKQPTGEQNAKKGVTDALSLGMNGLSQLLGGNGGLGGGQGNAGTGLDGSSL	240
QY	275	GGKGLQNLSPVDYQQLGNNAVGTGIGMKAGIQALNDIGTHRHSSTRSFVNKGDRAMAKEI	334
Db	241	GGKGLQNLSPVDYQQLGNNAVGTGIGMKAGIQALNDIGTHRHSSTRSFVNKGDRAMAKEI	300
QY	335	QGFMDQYPEVFGKPOYQKPGQEVKTDKSWAKALSKPDDGWTGPASMEQFNKAKGMIKR	394
Db	301	QGFMDQYPEVFGKPOYQKPGQEVKTDKSWAKALSKPDDGWTGPASMEQFNKAKGMIKR	360
QY	395	PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 437	
Db	361	PMAGDTGNGNLQARGAGSSSLGIDAMMAGDAINNMALGKLGA 403	
DR	RESULT 11		
XX	AAW06598		
XX	ID AAW06598 standard; Protein; 385 AA.		
XX	AC AAW06598;		
XX			


```

QY 215 QGSSCGKQPTGEQONAYKGVTDALSLGMNGLSQLLNGGLGGGQGNAGTGLDGSSL 274
DB 181 QGSSGKQPTGEQONAYKGVTDALSLGMNGLSQLLNGGLGGGQGNAGTGLDGSSL 240
QY 275 GKGGLNLSPVDYQOOLGNVAGTGIGMKAGTQALNDIGTHRHSSSTRSFVNKGDRAKAI 334
DB 241 GKGGLNLSPVDYQOOLGNVAGTGIGMKAGTQALNDIGTHRHSSSTRSFVNKGDRAKAI 300
QY 335 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMIKR 394
DB 301 GQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMIKR 360
QY 395 PMAGDTGNGNL 405
DB 361 PMAGDTGNGNL 371

RESULT 13
AAW06597
ID AAW06597 standard; Protein; 338 AA.
XX
AC AAW06597;
XX
DT 30-MAR-1997 (first entry)
DE
DE A hypersensitive response elicitor protein.
KW Hypersensitive response; elicitor; Erwinia chrysanthemi; plant;
KW disease-resistance; Escherichia coli; infiltration; virus;
KW bacterium; fungus; pathogen; biological control agent.
XX
OS Erwinia chrysanthemi.
XX
PN WO9639802-A1.
XX
PD 19-DEC-1996.
XX
PF 05-JUN-1996; 96WO-US08819.
XX
PR 07-JUN-1995; 95US-0475775.
XX
PA (CORR ) CORNELL RES FOUND INC.
PI Beer SV, Wei Z;
XX
DR WPI: 1997-051614/05.
DR N-PSDB: AAY49313.
XX
PT Imparting pathogen resistance to plants - with hypersensitive
PT response elicitor polypeptide or protein
XX
PS Claim 4; Page 44; 69pp; English.
XX
CC This sequence represents a hypersensitive response elicitor from
CC Erwinia chrysanthemi, with a mol.wt. of 34,000. The elicitor is
CC thermostable, has a glycine content of over 16%, and has no cysteine.
CC The elicitor may be used in a new method for imparting pathogen
CC resistance to plants, by application of the elicitor in a non-
CC infectious form to plant cells, by spraying, injection, leaf
CC abrasion, or plant infection with recombinant bacteria (non-
CC infectious to the host plant, e.g. Escherichia coli) expressing the
CC elicitor as a biological control agent, to allow recombinant
CC protein infiltration into the plant. The method confers virus,
CC bacterium or fungus disease-resistance on crops and ornamental
CC plants.
XX
SQ Sequence 338 AA;
Query Match 32.0%; Score 718.5; DB 18; Length 338;
Best Local Similarity 42.8%; Pred. No. 1.9e-48;
Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;
QY 47 MQTISI-CGAGGNNGLLGTSTRQNAGLGG-NSA---LGLGGGNQNDTVNLQAGLLTGMMMM 101

```

```

DB 1 MQITIKAHITGCDLGVSGLGAQ--GLKGLNSAASSLGSVDKLSSTIDKLTSALTSMM--- 55
QY 102 SMWGGGLMGGLGGLGNGLGSGGLGELSNALNDMLGGLSLTLGSKGNNTTSTTNS 161
DB 56 -----FGGALAQGLGAS-SKGLGMSNQLGQSFGN-----GAQASNLLSVPK- 96
QY 162 PLDQALGINSTSDNSTSDSTSDPMQQLLKMFEIMQSLFG-----DCQDQ 213
DB 97 -----SGGDALS-----KMFDAKLDLLGHDTVTTLTQNSQ 128
QY 214 TQSSSGGKQPTGEQONAYKGVTDALSLGMNGLSQLLNGGLGGGQGNAGTGLDGSS 273
DB 129 LANSMLNASQMTQGNMNAFGSGVNNALSSILGNLQOSM-----SGFSQPS 174
QY 274 LGGKGLNLSPVDYQOOLGNVAGTGIGMKAGTQALNDIGTHRHSSSTRSFVNKGDRAKAI 333
DB 175 LGAGGLGSLGAGAFNQLGNALGMVGQNAALSALSNVSTHVDGNRHFVDKEDRMAKE 234
QY 334 IGOFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTPASMEQFNKAKGMIK 393
DB 235 IGOFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDGDMTGASMDKFRQAMGMIK 294
QY 394 RPMAGDTGNGNLQARGAGGSLGIDAMMAGDAIINNMLKGLGAA 437
DB 295 SAVAGDTGNTNLNRGAGGASLGIDAAYVGDKIANNLSLGLANA 338

RESULT 14
AAW87638
ID AAW87638 standard; Protein; 338 AA.
XX
AC AAW87638;
XX
DT 09-MAR-1999 (first entry)
DE
DE A hypersensitive response elicitor protein.
KW Hypersensitive response elicitor protein; hairpin protein;
KW disease resistance; seed quality; insect control; corn borer;
KW Lepidoptera larvae; transgenic plant.
XX
OS Erwinia chrysanthemi.
XX
PN WO9854214-A2.
XX
PD 03-DEC-1998.
XX
PF 28-MAY-1998; 98WO-US10874.
XX
PR 30-MAY-1997; 97US-0048109.
XX
PA (CORR ) CORNELL RES FOUND INC.
PA (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI Beer SV, Laby RJ, Wei Z;
XX
DR WPI: 1999-070210/06.
DR N-PSDB: AAV83988.
XX
PT New fragments of an Erwinia hypersensitive response elicitor protein
PT and related DNA - used to impart disease resistance to plants, to
PT increase their growth and to control insects
XX
PS Disclosure; Page 7-8; 94pp; English.
XX
CC The present sequence represents a hypersensitive response elicitor
CC protein (also called hairpin protein) that is able to elicit a
CC hypersensitive response in plants. The specification also describes
CC hypersensitive response elicitors from other pathogenic organisms.
CC The protein, in non-infectious form, is applied to plants to impart
CC disease resistance (to a wide range of viral, bacterial and fungal
CC pathogens), to improve growth (yield, quantity and quality of seeds,

```

CC to provide earlier germination etc.) and to control insects (e.g. corn
 CC borers, Lepidoptera larvae etc.) The same results are provided by
 CC transgenic plants expressing the protein.

XX
 SQ Sequence 338 AA:
 Query Match 32.0%; Score 718.5; DB 20; Leng-h 338;
 Best Local Similarity 42.8%; Pred. NO. 1.9e-48;
 Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;
 Qy 47 MOISI-GGAGGNNLLCTSRONAGLGG-NSA---LGLGGGNQNDTVNOLAGLLTGMMMM 101
 Db 1 MQITIKAHIGDLGVSLGAQ--GLKGLNSAASSLGSSVDKLSIDKLTSAITSM--- 55
 Qy 102 SMGGGGLMGGGLGGLGNGLCGGGCGEGLSNALNDMLGSLNTLGSKGGNNTTSTTNS 161
 Db 56 -----FGGALAOGLGAS--SKGLGMSNQLGQSGFN-----GAQGASNLLSVPK- 96
 Qy 162 PLDQALGINSTSQNDSTSGTSDTSDSDPQQLLKMFSEIMQSLFG-----DQDQG 213
 Db 97 -----SGGDALS-----KMFDAKALDDLGLHDTVTKLITNOSNQ 128
 Qy 214 TOGSSSGGKQPTGEQNAKKGYTDALSGLMGNGLSQLLGGNGGGGAGTGLDGSS 273
 Db 129 LANSMLNASQMTQGNMNAFGSVNNALSSILGNGLGQSM-----SGFSQPS 174
 Qy 274 LGCKGLONLSCPVDYQOLGNVCTGICMKAGTQALNDICTHRHSSTRSFVNKGDRAKMA 333
 Db 175 LGAGGLGSLGAGAFNOLGNAIGMGVGNAAUSLSNVSTHYDGNRRHFVDKEDRGMAK 234
 Qy 334 IGQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDDDGMPASMEQFNKAKGM 393
 Db 235 IGQFMDQYPEIFGKPEYOKDGNSSPKTDDKSWAKALSKPDDDDGMTGASMDKFRQAMGM 294
 Qy 394 RPMAGDTGNGNLQARGAGSSGLGIDAMMAGDAINNMALGKLGA 437
 Db 295 SAVAGDTGNTNLRGAGGASLGIDAAVVGDKIANNSLGKLANA 338

RESULT 15

AAW82407
 ID AAW82407 standard; Protein; 338 AA.

AC AAW82407;

DT 23-FEB-1999 (first entry)

DE E. chrysanthemi HrpN-ECH protein.

XX Hypersensitive response elicitor protein; hrpN-Ech; pathogen resistance;
 KW plant; transformation; pathogen-inducible promoter.

XX Erwinia chrysanthemi.

XX U95850015-A.

PN 15-DEC-1998.

PF 07-JUN-1995; 95US-0484358.

PR 07-JUN-1995; 95US-0484358.

XX (CORR) CORNELL RES FOUND INC.

XX Bauer D, Collmer A;

PI WPI; 1999-069852/06.

DR N-PSDB; AAV73494, AAV73507.

XX DNA encoding Erwinia chrysanthemi hypersensitive response protein

PT hrpN - useful for imparting pathogen resistance to plants

XX Claim 2; Column 29-30; 27pp; English.

PS

XX This sequence represents a novel Erwinia chrysanthemi protein, hrpN-Ech,
 CC that elicits a hypersensitive response in plants. The encoding DNA can be
 CC used for imparting pathogen resistance to plants, by transforming a
 CC plant with a vector containing the DNA and a pathogen-inducible promoter.

XX Sequence 338 AA:

Query Match 32.0%; Score 718.5; DB 20; Length 338;
 Best Local Similarity 42.8%; Pred. NO. 1.9e-48;
 Matches 173; Conservative 41; Mismatches 111; Indels 79; Gaps 11;
 Qy 47 MOISI-GGAGGNNLLCTSRONAGLGG-NSA---LGLGGGNQNDTVNOLAGLLTGMMMM 101
 Db 1 MQITIKAHIGDLGVSLGAQ--GLKGLNSAASSLGSSVDKLSIDKLTSAITSM--- 55
 Qy 102 SMGGGGLMGGGLGGLGNGLCGGGCGEGLSNALNDMLGSLNTLGSKGGNNTTSTTNS 161
 Db 56 -----FGGALAOGLGAS--SKGLGMSNQLGQSGFN-----GAQGASNLLSVPK- 96
 Qy 162 PLDQALGINSTSQNDSTSGTSDTSDSDPQQLLKMFSEIMQSLFG-----DQDQG 213
 Db 97 -----SGGDALS-----KMFDAKALDDLGLHDTVTKLITNOSNQ 128
 Qy 214 TOGSSSGGKQPTGEQNAKKGYTDALSGLMGNGLSQLLGGNGGGGAGTGLDGSS 273
 Db 129 LANSMLNASQMTQGNMNAFGSVNNALSSILGNGLGQSM-----SGFSQPS 174
 Qy 274 LGCKGLONLSCPVDYQOLGNVCTGICMKAGTQALNDICTHRHSSTRSFVNKGDRAKMA 333
 Db 175 LGAGGLGSLGAGAFNOLGNAIGMGVGNAAUSLSNVSTHYDGNRRHFVDKEDRGMAK 234
 Qy 334 IGQFMDQYPEVFGKPYQKPGQEVKTDKSWAKALSKPDDDDGMPASMEQFNKAKGM 393
 Db 235 IGQFMDQYPEIFGKPEYOKDGNSSPKTDDKSWAKALSKPDDDDGMTGASMDKFRQAMGM 294
 Qy 394 RPMAGDTGNGNLQARGAGSSGLGIDAMMAGDAINNMALGKLGA 437
 Db 295 SAVAGDTGNTNLRGAGGASLGIDAAVVGDKIANNSLGKLANA 338

Search completed: January 22, 2003, 16:28:24

Job time : 42 secs